RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT

VOLUME 4

DAVID J. KETCHEN, JR.
DONALD D. BERGH

Editors
RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT
RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT

Series Editors: David J. Ketchen, Jr. and Donald D. Bergh
CONTENTS

LIST OF CONTRIBUTORS ix

INTRODUCTION xi

SPECIAL SECTION ON METHODS AND THE RESOURCE-BASED VIEW OF THE FIRM

TEN YEARS AFTER: SOME SUGGESTIONS FOR FUTURE RESOURCE-BASED VIEW RESEARCH
   Urs S. Daellenbach and Michael J. Rouse 3

RESOURCE COMPLEMENTARITY, INSTITUTIONAL COMPATIBILITY, AND SOME ENSUING METHODOLOGICAL ISSUES IN TESTING THE RESOURCE-BASED VIEW
   Tailan Chi and Edward Levitas 19

MIXED METHODS IN STRATEGY RESEARCH: APPLICATIONS AND IMPLICATIONS IN THE RESOURCE-BASED VIEW
   José F. Molina-Azorín 37

TOWARD GREATER INTEGRATION OF THE RESOURCE-BASED VIEW AND STRATEGIC GROUPS RESEARCH: AN ILLUSTRATION USING RANDOM COEFFICIENTS MODELING
   Jeremy C. Short 75
SPECIAL SECTION ON METHODS IN INTERNATIONAL STRATEGY RESEARCH

MODELING INTERNATIONAL EXPANSION
Xavier Martin, Anand Swaminathan and Laszlo Tihanyi 103

STATIC TRIANGULAR SIMULATION AS A METHODOLOGY FOR INTERNATIONAL STRATEGIC MANAGEMENT RESEARCH
George S. Yip, G. Tomas M. Hult and Audrey J. M. Bink 121

NEW METHODS FOR EX POST EVALUATION OF REGIONAL GROUPING SCHEMES IN INTERNATIONAL BUSINESS RESEARCH: A SIMULATED ANNEALING APPROACH
Paul M. Vaaler, Ruth V. Aguilera and Ricardo Flores 161

ADDITIONAL EXEMPLARY CONTRIBUTIONS

APPLYING ADVANCED PANEL METHODS TO STRATEGIC MANAGEMENT RESEARCH: A TUTORIAL
Peter Hom and Katalin Takacs Haynes 193

INTERPRETING EMPIRICAL RESULTS IN STRATEGY AND MANAGEMENT RESEARCH
J. Myles Shaver 273

MEDIATION IN STRATEGIC MANAGEMENT RESEARCH: CONCEPTUAL BEGINNINGS, CURRENT APPLICATION, AND FUTURE RECOMMENDATIONS
Toyah L. Miller, María del Carmen Triana, Christopher R. Reutzel and S. Trevis Certo 295
Contents

USING POLICY CAPTURING TO UNDERSTAND STRATEGIC DECISIONS – CONCEPTS AND A MERGERS AND ACQUISITIONS APPLICATION
Amy L. Pablo 319

TESTING ORGANIZATIONAL ECONOMICS THEORIES OF VERTICAL INTEGRATION
Kaouther Lajili, Marko Madunic and Joseph T. Mahoney 343

AN ONTOLOGICAL FOUNDATION FOR STRATEGIC MANAGEMENT RESEARCH: THE ROLE OF NARRATIVE
Dominik Heil and Louise Whittaker 369
This page intentionally left blank
# LIST OF CONTRIBUTORS

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruth V. Aguilera</td>
<td>Department of Business Administration, College of Business, Champaign, IL, USA</td>
</tr>
<tr>
<td>Audrey J. M. Bink</td>
<td>Uxbridge College, Uxbridge, Middlesex, UK</td>
</tr>
<tr>
<td>S. Trevis Certo</td>
<td>Texas A&amp;M University, Mays Business School, College Station, TX, USA</td>
</tr>
<tr>
<td>Tailan Chi</td>
<td>University of Kansas, School of Business, Lawrence, KS, USA</td>
</tr>
<tr>
<td>Urs S. Daellenbach</td>
<td>Victoria Management School, Victoria University of Wellington, Wellington, New Zealand</td>
</tr>
<tr>
<td>Ricardo Flores</td>
<td>Department of Business Administration, College of Business, Champaign, IL, USA</td>
</tr>
<tr>
<td>Katalin Takacs Haynes</td>
<td>Department of Management, Texas A&amp;M University, Mays Business School, College Station, TX, USA</td>
</tr>
<tr>
<td>Dominik Heil</td>
<td>Wits Business School, University of the Witwatersrand, Wits, South Africa</td>
</tr>
<tr>
<td>Peter Hom</td>
<td>Department of Management, W. P. Carey School of Business, Arizona State University, Tempe, AZ, USA</td>
</tr>
<tr>
<td>G. Tomas M. Hult</td>
<td>The Eli Broad Graduate School of Management, Michigan State University, East Lansing, MI, USA</td>
</tr>
<tr>
<td>Kaouthar Lajili</td>
<td>Telfer School of Management, University of Ottawa, Ottawa, ON, Canada</td>
</tr>
<tr>
<td>Edward Levitas</td>
<td>University of Wisconsin-Milwaukee, Sheldon B. Lubar School of Business Administration, Milwaukee, WI, USA</td>
</tr>
</tbody>
</table>
LIST OF CONTRIBUTORS

Marko Madunic
Department of Business Administration, College of Business, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Joseph T. Mahoney
Department of Business Administration, College of Business, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Xavier Martin
Tilburg University, Faculty of Economics and Business Administration and Center for Economic Research, Tilburg University, Tilburg, The Netherlands

Toyah L. Miller
Texas A&M University, Mays Business School, College Station, TX, USA

José F. Molina-Azorín
University of Alicante, Depart. Organizacion Empresas, Alicante, Spain

Amy L. Pablo
University of Calgary, Haskayne School of Business, Calgary, Alberta, Canada

Christopher R. Reutzel
Texas A&M University, Mays Business School, College Station, TX, USA

Michael J. Rouse
Richard Ivey School of Business, University of Western Ontario, London, ON, Canada

J. Myles Shaver
Carlson School of Management, University of Minnesota, Minneapolis, MN, USA

Jeremy C. Short
Texas Tech University, Rawls College of Business Administration, Lubbock, TX, USA

Anand Swaminathan
Goizueta Business School, Emory University, Atlanta, GA, USA

Laszlo Tihanyi
Texas A&M University, Mays Business School, College Station, TX, USA

María del Carmen Triana
Texas A&M University, Mays Business School, College Station, TX, USA

Paul M. Vaaler
Department of Business Administration, College of Business, Champaign, IL, USA

Louise Whittaker
Wits Business School, University of the Witwatersrand, Wits, South Africa

George S. Yip
Capgemini Consulting, London Business School and AIM Research, UK
INTRODUCTION

Welcome to the fourth volume of Research Methodology in Strategy and Management (RMSM). The publication of our fourth volume provides a source of satisfaction because our original contract with Elsevier only guaranteed a three volume run for the series. The popularity of RMSM led our contacts at the publisher to be eager to continue beyond their original commitment. We are excited about the future of the series, and have begun assembling Volume 5.

Before we discuss the contents of this volume, let us briefly reflect on its predecessor. Volume 3 was debuted at the 2006 Academy of Management meeting in Atlanta. Just like with the first two volumes, Volume 3 was showcased in a symposium sponsored by the Business Policy and Strategy division. We were thrilled that our meeting room was filled to capacity. We believe this turnout reflects a strong desire by strategy researchers to improve their methodology skills; a desire we hope this book series is serving capably. We want to thank Joel Baum, Mason Carpenter, Teppo Felin, Nicolai Foss, Bill McKelvey, Nathan Podsakoff, Phil Podsakoff, Gregory Reilly, and Wei Shen for offering excellent presentations of their chapters.

We hope that Volume 4 also will be well received. The volume you hold in your hands offers 13 chapters. The first four chapters comprise a special section on “Methods and the Resource-Based View of the Firm.” One of the highlights of Volume 2 of this series was a piece on testing resource-based theory by Jay Barney and Tyson Mackey. Because this chapter was so well received, we recruited a diverse and skilled set of authors to provide detailed discussions on various aspects of research into this prominent theory. In the first chapter, Urs Daellenbach and Michael Rouse extend their earlier discussion of testing the theory that was offered in the Strategic Management Journal. They offer a series of insightful recommendations designed to inform future inquiry.

Chi and Levitas react to the recent suggestion that value is exogenous to the firm; specifically that value is driven by external forces rather than by internal firm mechanisms. They argue this is a theoretical misconception that can lead to biased and inconsistent parameters estimate when attempting to
empirically verify theoretical components of the resource-based view (RBV). In particular, they suggest that viewing resource value as exogenous oversimplifies the context within which firms operate. In the third part of the special section, José F. Molina-Azorín examines how to combine the strengths of quantitative versus qualitative methods each in mixed methods studies of the RBV. Finally, Jeremy Short lays out an agenda for integrating RBV research with strategic groups research. Scholars have called for such integration since the late 1980s, but it has been slow in forthcoming. Short’s hope is to foster integrative approaches that will shed new light on why some firms outperform others. Taken together, these four chapters build effectively on the effort of Barney and Mackey, and they provide a wealth of ideas for scholars interested in further developing the RBV.

The next three chapters compose a special section on “Methods in International Strategy Research.” Research that steps beyond domestic borders confronts unique challenges. We were fortunate to attract three skilled authors teams to discuss some of the challenges. First, Xavier Martin, Anand Swaminathan, and Laszlo Tihanyi address how to model discrete strategic decisions, such as choices about location and mode of entry. Next, George S. Yip, G. Tomas M. Hult, and Audrey Bink present “static triangular simulation” as a method with the potential to assess some hard-to-measure phenomena in the international arena including global relationship management. In the third component of the special section, Paul Vaaler, Ruth Aguilera, and Ricardo Flores present an innovative method for identifying regional groups of countries. Their chapter promises to be a popular one because regional groups have become a central concern in research journals as well as in executive suites.

The remainder of this volume contains a series of quality efforts from renowned and emerging scholars. The first two chapters take significant steps toward helping us to resolve vexing issues. Peter Hom and Katalin Haynes provide a detailed tutorial on analyzing panel data. Examining this kind of data has proved daunting for a long time, so the clarity Hom and Haynes offer will be welcomed by their colleagues. Next, Myles Shaver argues and demonstrates that researchers should place emphasize on interpreting the magnitude of coefficient estimates rather than only assessing statistical significance. In particular, Shaver shows that an exclusive focus on significance can lead to erroneous conclusions.

Next, Toyah Miller, María del Carmen Triana, Christopher Reutzel, and Trevis Certo assess how mediation analyses are used in strategy research. They make it clear that some longstanding practices need to be revisited. As with several of our past chapters, this chapter reflects collaboration
between a professor and a team of doctoral students. While we are always on
the lookout for good chapters, we take special pride in publishing chapters
that help doctoral students get started as researchers.

Amy L. Pablo offers an excellent primer on how to use policy capturing to
understand strategic decisions. Calls continue for more inquiry into how
executives form strategy. Through this book series, we have attempted to
highlight opportunities to do so. Previous volumes have included chapters
on repertory grid technique, cause mapping, and conjoint analysis. Pablo’s
chapter offers an effective complement to these past pieces. We believe that
they collectively offer a diverse toolbox for those scholars interested in
tapping into strategy process.

Vertical integration has long been and continues to be an important topic
of strategic management research. Kaouthar Lajili, Marko Madunic, and
Joseph Mahoney capture the theoretical and empirical contributions from
theories of organizational economics (especially) for vertical integration,
as well as for vertical contracting. Particular attention is devoted to
transaction cost economics and agency theory. Mahoney is uniquely
qualified to comment on the intersection of strategy and economics, thus
we believe the thoughts on future research that this chapter offers are
worthy of careful attention.

Our final chapter offers, we think, a special treat. Although we have
had a few authors from Europe and Asia, the series has been dominated by
United States based authors. Meanwhile, both the series and our own
research have been largely positivist in orientation. The final chapter breaks
both of these molds. Dominik Heil and Louise Whittaker, both based in
South Africa, offer a decidedly non-positivist approach to the strategic
management field. Their chapter is among the most thought provoking
included in the series to date, and we hope you enjoy reading an
unconventional and insightful take on the strategy field.

Overall, we hope you enjoy this collection. We are very grateful to all of
the contributors for their wisdom and efforts.

David J. Ketchen, Jr.
Donald D. Bergh
Editors
This page intentionally left blank
SPECIAL SECTION ON METHODS AND THE RESOURCE-BASED VIEW OF THE FIRM
TEN YEARS AFTER: SOME SUGGESTIONS FOR FUTURE RESOURCE-BASED VIEW RESEARCH

Urs S. Daellenbach and Michael J. Rouse

ABSTRACT

While research related to the resource-based view (RBV) has expanded markedly in the last decade, debates continue over the theory, the extent to which our understanding of the theory has been advanced in a meaningful way, and the most appropriate approaches for empirical RBV research. We present some additional perspectives on current debates, summarize key challenges that empirical studies face, and offer some suggestions and directions for future RBV research.

INTRODUCTION

In over 10 years since we first began presenting our arguments about research methods for the resource-based view (RBV) (published later as Rouse & Daellenbach, 1999), it is a little surprising to see that at times the debate has not altered that markedly. For example, most recently, Arend (2006, p. 412) critiques the RBV empirical literature (60 studies), stating that while “[a]ll provided significant support for their hypotheses related to the
RBV [...] none of them provides an adequate test of the RBV,” with the failure to measure performance-related aspects being the most troubling feature. Thus, while the popularity of resource-based research is testament to the sentiment that it constitutes a flagship theory of strategic management, Arend (2006, p. 418) argues that this may be a “false halo” lacking appropriate empirical support.

Why has such empirical support not been forthcoming or less critically, why does it continue to be unconvincing? We think that a core reason may have to do with the specific nature of the RBV as a theory which, in turn, requires a specific methodology for empirical testing. Intuitively, the RBV is appealing – it just seems to make sense that heterogeneous resource endowments, some of which are rare and inelastic in supply, should lead to superior organizational performance. We propose that a core requirement for empirically testing the RBV is measurement and analysis at the resource level. This means that RBV researchers require a quality and quantity of longitudinal data that generally are not, or have not been, easily available. Linked directly to the issue of resource-level measurement is the requirement for a comparable dependent variable. Performance is generally aggregated at the firm or business-unit level, often reflecting multiple strategic thrusts. Since different strategies involve the pursuit of different advantages, it would seem to be important that a level of performance specific to a particular strategy also be included, one which can be associated more directly with the deployment of a resource or resource bundles. Thus, we augment the arguments presented by Barney and Mackey (2005) in the previous volume of this series by fully highlighting the benefits for RBV research from using a more disaggregated dependent variable.

Our discussion begins by noting that more care needs to be taken to link the methodology to the theory being advanced. To illustrate the data collection and sample selection considerations associated with RBV theory, we look specifically at the situations typically faced by firms when implementing strategies with heterogeneous, inelastic resources. We follow this with suggestions for future directions in RBV research, considering both the direction underlying most current RBV research as well as how other RBV attributes have received less attention. None of this is to suggest that progress has not been made by studies to date. Rather, we argue that access to data continues to be the biggest barrier to RBV research and it is to the credit of the strategy discipline that, notwithstanding the data barrier, significant testing has in fact occurred (Barney & Arikan, 2001). As noted previously, the nature of the RBV theory and propositions related to sustained competitive advantage demand special consideration of the
research framework employed, with particular focus on sample selection and data collection (Rouse & Daellenbach, 1999). Fundamentally, as with other theoretical perspectives, there is a need for a match between the theory and the research methodology. Thus, we start by revisiting key aspects of the RBV to highlight implications for the methods employed. These aspects and the research design implications are summarized in Table 1.

THE RBV AND COMPETITIVE (DIS)ADVANTAGE

The RBV “asserts that firms gain and sustain competitive advantages by deploying valuable resources that are inelastic in supply” (Ray, Barney, & Muhanna, 2004, p. 23). More specifically, Barney (1991) argues that for sustained advantage, resources need to exhibit four key attributes – VRIS: value, rareness, imperfect imitability and (lack of) substitutability. Powell (2001) augments this in noting that firms may simultaneously have competitive advantages and disadvantages, as do West and deCastro (2001) when emphasizing the inclusion of resource weaknesses and distinctive inadequacies to add to the explanatory and predictive power of theories related to competitive advantage. Yet, in attempting to explain superior performance, strategy research somewhat naturally has focussed primarily on identifying sources of competitive advantage. This focus is longstanding in strategy research where concepts such as the match between industry key success factors and firm strengths (e.g., de Vasconcellos & Hambrick, 1989) or distinctive competence (e.g., Snow & Hrebiniak, 1980) are hypothesized as linearly related to performance.

Table 1. Key Aspects of the RBV and their Implications for Research Design.

<table>
<thead>
<tr>
<th>Assumptions about Firms Underlying the RBV</th>
<th>Research Design Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firms have areas of competitive advantage, parity, and disadvantage.</td>
<td>Overall firm performance results from the effects of both relative strengths and weaknesses.</td>
</tr>
<tr>
<td>Firms deploy bundles of resources/ capabilities.</td>
<td>Broader sets of resources need to be analyzed.</td>
</tr>
<tr>
<td>Advantages/disadvantages are relative to other firms pursuing similar value creation activities.</td>
<td>Samples must be selected to include firms seeking comparable value creation activities.</td>
</tr>
</tbody>
</table>
In contrast, others have considered contingencies in these relationships by categorizing variables as success producers and/or failure preventers (Varadarajan, 1985), or modelling differing effects depending on industry stage (Powell, 1992) or context (Hitt & Ireland, 1985; Miller & Shamsie, 1996).

When considering the effects of the deployment of resources/capabilities, it is likely that while some of a firm’s resources may produce an advantage, others will lead to either competitive parity or disadvantage. This suggests that both areas of relative strength and weakness need to be assessed when attempting to identify sources of sustained advantage.

However, due to the difficulty in fully specifying and/or measuring all of a firm’s resources (Godfrey & Hill, 1995; Poppo & Weigelt, 2000), studies may focus only on resources that appear to separate out the better-performing firms (Arend, 2006), assuming that other resources are generic to all firms (i.e., that such resources may be no more than basic requirements for being a player in a particular industry) and hence have no effect (positive or negative) on above-average performance. One example of such an approach is Ainuddin, Beamish, Hulland, and Rouse (2007), who utilized the context of international joint ventures to assess whether a particular set of resources (those with VRIS characteristics) affect performance. Their study of 96 Malaysian international joint ventures using survey and performance data provides empirical evidence that supports the RBV assertion that resources characterized by VRIS attributes have an influence on – in their study, joint venture – performance.

Additional arguments for gathering data on a broader set of resources can be seen in cases where firms create value and exhibit high performance yet appear to have few rare, inimitable or non-substitutable resources (see, e.g., Barney’s 1997, p. 156, discussion of The Mailbox, Inc. where “success depends not on the ability to do one thing well, but rather on the ability to do the hundreds of thousands of things needed to run a bulk mailing business well.”). Examples of this kind indicate that it can be crucial to consider all sources of advantage, parity and disadvantage simultaneously. Without such consideration, it is possible that some studies may be over- or under-estimating the impact of the particular resources measured because important explanatory variables have not been included in the analyses.

The argument that firms possess and typically deploy bundles of resources, only some of which may lead to a competitive advantage and associated above-average performance, is not new (Penrose, 1959; Black & Boal, 1994; Barney, 1997; Lippman & Rumelt, 2003). However, because untangling the complexity of tangible and intangible resource bundles
and the resulting value created poses additional measurement difficulties (Godfrey & Hill, 1995), much RBV research has focussed on the effects of individual, separable and identifiable resources or have labelled more complex resource combinations as capabilities without assessing specifically the resources that combine to create them (e.g., Makadok & Walker, 2000). This tendency to identify only individual resources as satisfying the attributes of the RBV is potentially limiting the findings and contribution made by empirical research and may be partly responsible for the continuing critique that the RBV receives.

This discussion suggests an interesting, and unresolved issue in RBV theory – does competitive advantage accrue primarily from resources or from the combination of resources and capabilities together? Barney (1997) suggests that both are important by reformulating VRIS (which refers to resource attributes) as VRIO, i.e., value, rarity, inimitability/non-substitutability, and organization, where VRI captures the VRIS attributes, and organization is cited as the key implementation/organizing element. That is, organization is a necessary condition for generating above normal performance from VRIS resources that, by definition, hold the potential for generating performance advantages. Organizing becomes, therefore, a capability (composed of still more capabilities and resources) for leveraging firms’ resources or, as articulated by Barney and Mackey (2005), to create and implement strategies. Barney’s (1997) discussion of The Mailbox, Inc., suggests that implementation (organization for leveraging resources) may itself be a capability. Clearly, there are elements of organization that likely cannot be sources of advantage (Barney & Mackey, 2005) precisely because organizational structures, for example, can easily be imitated. Structures of relationships embedded within organizational structures could, however, be sources of advantage due to their inherent inimitability based on social complexity, but if and only if those structures can be linked to value generation, and are non-substitutable by other (perhaps similar or equivalently functioning) structures. In combination, the points above further indicate the need for a level of detail in data on firms and industries when empirically testing the RBV that is not present in the majority of strategy data sets (Hitt, Boyd, & Li, 2004).

The notion that bundles of resources are purposefully deployed in pursuit of competitive advantage from a particular strategy does, though, inherently provide useful guidance for data collection. Specifically, it implies that RBV research can benefit from focusing on the distinctive value that groups of firms are attempting to create. Following this rationale, Ray et al. (2004) propose and use a measure of the effectiveness of business processes (which, for their insurance company study, is customer satisfaction) as a
proxy of the value generated as did Henderson and Cockburn (1994) who use research productivity in pharmaceutical firms. Dutta, Narasimhan, and Rajiv (2005) extend this by designing a methodology to use output measures to infer capabilities across firms, again drawing on more fine-grained indicators of the outcome of deployment of resources in a firm’s strategy rather than overall firm performance. Similarly, Lieberman and Dhawan (2005) use stochastic production frontiers to develop measures of resources and capabilities and test linkages to performance. These developments suggest that some RBV researchers have been moving forward in measuring resources and their value, although they have had to craft their analyses through richer measurement of value and not simply overall firm performance.

The second aspect highlighted above is the need to compare the value created across firms pursuing similar strategies. The relative dimension of the term ‘advantage’ has often been overlooked in empirical RBV research, with studies aggregating data across firms in an industry simply to generate sufficient sample size for their chosen statistical techniques rather than to allow meaningful comparisons of advantage, parity and disadvantage. For example, Richard (2000) draws on the RBV in considering the relationships among racial diversity, business strategy and firm performance in the US banking industry, yet uses only dummy variables to assess all differences in the relationships that might exist in the four US states covered by the sample. Similarly, Schroeder, Bates, and Juntila (2002) use a stratified sample of manufacturing plants across three industries and five countries. Neither study indicates whether the samples included at least some firms with competitive advantages (which is likely), and so it is feasible that the resources measured do not characterize those distinguishing firms with sustained advantages and that the results reflect primarily average performance levels.

In another example, Santhanam and Hartono (2003), replicating and extending Bharadwaj’s (2000) study on IT capability by controlling for industry average and prior firm performance, selected their sample purely from a ranking of top IT firms across industries. Thus, the association between higher performance levels and IT is potentially spurious as numerous other resources that may be valuable across the SIC codes were not included. The differences in their results between the 2-digit and 4-digit SIC samples may also indicate that testing for context-specific relationships was warranted. The potential finding that an IT capability is a valuable resource only in some circumstances would not be overly surprising, but was not explicitly considered in the analyses. Thus, while some theories and
statistical techniques benefit from random samples (Hitt et al., 2004), the RBV is explicitly a relative perspective that contrasts advantages and disadvantages across specific firms and so there is a need to ensure that the research methodology fits the RBV theory being tested.

Some studies have, though, included this consideration. Henderson and Cockburn’s (1994) “architectural competence” was found to be valuable in pharmaceutical firms engaged in the discovery and marketing of new drugs. It may, though, be irrelevant for pharmaceutical firms pursuing different strategies, for example, generic drug manufacturers for whom patents are a mobility barrier and do not provide them with competitive advantage in their strategic space. When samples cut across strategies (strategic groups) or industries, therefore, the relationship between resources and value or performance may be masked since some resources may not be crucial to generate competitive advantages for all strategies. Capabilities associated with adding value through customer service, for example, may be less crucial for firms using lower prices and efficient operations to achieve a low-cost advantage. Market commonality, as emphasized alongside resource similarity in Chen’s (1996) framework identifying the extent of competitive rivalry, may serve as another key sample selection criterion if competitive advantages and the value created by key bundles of resources can be expected to vary across markets. In other instances, though, where similar strategies are replicated across markets that are defined geographically, collecting rich primary data on value creation and resources longitudinally may provide the sample size necessary to estimate quantitatively the impact of particular resources, capabilities and bundles. For example, even though Southwest, Ryanair, Westjet and other leading regional low-cost airlines do not compete directly, with suitable data the contribution of specific resources and bundles associated with this strategy could be more definitively identified.

While the discussion above reiterates and expands on the arguments we presented relating to careful sample selection (Rouse & Daellenbach, 1999), the final step in the proposed method focussed on the need for rich data. While we argued for the use of comparative, in depth case studies and indicated how an ethnographic approach may hold benefits, it is likely that a set of ‘quantitative case studies’ (Barney & Mackey, 2005) could be equally useful, providing potentially broader possibilities for generalizing the results. Studies of the RBV cannot rely purely on secondary sources as such data are unlikely to adequately measure the resource bundles being deployed by firms or to quantify the distinctive value that firms in a suitable sample are attempting to generate. However, the arguments above that
conclude that a non-random sample is needed to test the RBV, also mean that it is likely that this sample will be more limited in number relative to those generated for large-scale cross-sectional research and may make the collection of detailed RBV data more feasible.

VALUE CREATION THROUGH RBV RESEARCH

The final point above raises the question of the extent to which we should be seeking broad generalizability from empirical RBV research. As noted, there are some situations where the results of empirical analyses may be generalized across a range of settings (e.g., across countries/regions for the resources associated with the value creating low-cost airline strategy). In other circumstances, though, the value of a resource and its ability to support sustained advantage may be restricted to a particular strategic group, industry or market. The notion that the value of a resource may be context specific is not particularly surprising or problematic. One need look no further than the extensive mergers and acquisitions literature for examples of the failed pursuit of synergies (see Larsson & Finkelstein, 1999, for a synthesis of previous findings), often where it was assumed that resources and capabilities that had proven valuable in one context would prove equally valuable in another setting. Thus, we would argue that broad generalizability need not be the definitive test of the extent to which empirical RBV research adds value.

On the other hand, the growing literature testing the RBV has already identified a broad range of resources associated with sustained advantage and the settings in which this is likely to occur (see Barney & Arikan, 2001 for a review; or more recent studies including de Carolis, 2003; Lieberman & Dhawan, 2005; Ainuddin et al., 2007). Extending this list further by more fully specifying situations in which a resource generates a sustained advantage is likely of incremental value if pursued without consideration of sample selection and the type of data collected. To an extent, such studies are already appearing as other business disciplines become familiar with resource-based theory and apply it to their domain of interest (e.g., Lopez, 2003; Santhanam & Hartono, 2003; Wade & Hulland, 2004). If limited attention is paid to matching the research design to the underlying RBV assumptions, the insights concluded are not seen as adding distinctive value to the RBV and could probably have been as readily achieved through developing suitable teaching case studies or practitioner applications of RBV analyses as outlined in the growing list of strategy and management
textbooks that now include the RBV (e.g., Barney, 1997). On the other hand, if new RBV research can identify/specify resources not previously highlighted as contributing to advantages and disadvantages or contextualize the type and extent of value created, then these will help to extend the contribution of the RBV to the field of strategic management.

Beyond documenting the full range of situations where particular resources fulfil the VRIS questions set out in Barney (1991), what might be some suggestions for future empirical RBV research which could add more/distinctive value? What needs to be done to facilitate RBV research and to ground the theory empirically?

**SUGGESTIONS FOR FUTURE EMPirical RBV RESEARCH**

The following section highlights a range of potential avenues for future empirical RBV research. These are summarized in Table 2.

The majority of empirical RBV research to date has focussed on whether particular resources are associated with firm performance. Such studies have added value when they have carefully selected the firms’ analysed and considered performance longitudinally, so that a clear timeframe and context are established for the potential sustained competitive advantages. Where additional value can be added is through empirical studies, which make a direct connection between a specific resource and a measure of disaggregated performance. By disaggregated performance, we mean a performance outcome that is as closely as possible attributable to particular bundles of resources. This has occurred to an extent in some recent studies (e.g., Ray et al., 2004; Hult & Ketchen, 2001) and is consistent with the logic embedded in the popular management control framework the Balanced

<table>
<thead>
<tr>
<th>Table 2. Future Directions for RBV Research.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linking resource bundles to disaggregated performance outcomes as well as overall firm performance.</td>
</tr>
<tr>
<td>Assessing rent generation as well as rent capture, appropriation.</td>
</tr>
<tr>
<td>Simultaneously considering sources of competitive advantage, parity, and disadvantage, as well as their interrelationships.</td>
</tr>
<tr>
<td>For comparable groups of firms, contrasting performance levels to identify success producer and failure preventer resources.</td>
</tr>
<tr>
<td>Analyzing barriers to imitation, imitation processes, and substitution mechanisms.</td>
</tr>
</tbody>
</table>
Scorecard (Kaplan & Norton, 1992), where financial performance is modelled as the end result of a range of other value creation processes (such as customer retention, satisfaction, innovation, or employee productivity and motivation). Thus, multiple sources of value creation are explicitly modelled and firms are encouraged to measure these intermediating outcomes. RBV research could benefit from a similar approach in that superior value creation sustained over time can then be linked both to resource bundles as well as overall firm performance attributable to combined advantages and disadvantages.

The measurement of intermediating outcomes is also consistent with Coff’s (1999) argument about rent appropriation. He states that the RBV was formulated to identify the conditions under which firms may generate rents and does not specify who captures the rents, indicating that overall measures of firm performance may face difficulties when used as the dependent variable for RBV research. While sustained high performance is clearly indicative of some competitive advantages, sustained competitive advantages may also be present even when firm performance is at average levels if some stakeholders are appropriating the benefits of the value created.

Germane to this area are a range of resources that resource-based theory has highlighted as potentially contributing to advantage even though these resources are not solely controlled or owned by the organizations themselves. These resources include reputation (Dierickx & Cool, 1989; Deephouse, 2000), cooperative advantage (Lei, Slocum, & Pitts, 1997; Dyer & Singh, 1998; Combs & Ketchen, 1999) and customer relationships (Luo, Griffith, Liu, & Shi, 2004). For some of these, questions about rent appropriation are central. The unique characteristics of such resources should also alert researchers to assess the costs incurred with developing such resources. Reputation, while offering substantial potential advantages including the benefit of the doubt, can be particularly constraining on strategies in that consistency of actions is required to both build and maintain reputations. An irony is that investment in building and maintaining reputation (e.g., brand equity), may reduce performance, whereas spending brand equity without regard for long-term brand/reputation effects incurs lower investment and hence may yield better short-term performance. Furthermore, such resources as reputation or brand are cooperatively deployed and so firms cannot simply choose when and where they utilize them. Additional examination of this class of resources, thus, appears warranted.

At the resource level, misspecification may also be occurring in that most studies tend to consider only a subset of the resource bundles deployed
(Arend, 2006), rather than characterizing the full set that might impact performance and value creation. This broader consideration is necessary when samples cut across firms with high, average and low performance, with the value created by some resources potentially understated if areas of resource parity and weakness are not included in the analysis. West and DeCastro (2001) present a range of propositions for how strengths and weaknesses might be interrelated, providing another potential avenue for RBV research and augmenting Leonard-Barton’s (1995) study suggesting how core capabilities can become core rigidities.

Research that starts by stratifying performance more systematically within an industry or strategic group provides another alternative to current approaches. If resource bundles can then be measured in detail for low-, average- and/or high-performing firms, contrasts can be made between the performance levels to identify resources that are generic to all firms or distinguishing low from average/high or high from average. To identify such distinctions, Hansen, Perry, and Reese (2004, pp. 1282, 1283) propose the use of Bayesian methodology to more readily “focus on truly firm-specific phenomena” and avoid “an important lack of congruency between RBV theory and regression-type analysis.”

The cornerstones of advantage set out by Barney (1991) and Peteraf (1993) offer other research possibilities. Most studies of the RBV present coherent and logical arguments for how the particular resources of interest satisfy conditions allowing ex ante and ex post limits to competition, imperfect mobility and heterogeneity. However, the mechanisms through which these conditions are sustained have received less consideration (Priem & Butler, 2001). While intangible and immobile resources are acknowledged as a central concern of the RBV (Peteraf, 1993; Rouse & Daellenbach, 1999), resource-based theory holds additional promise in directing our attention to the processes associated with imitation and substitution, and why competitors are unable to duplicate the value created by those with advantages. For example, Woiceshyn and Daellenbach (2005) examine why an innovation to the oil industry, horizontal drilling, even though available to all and known by the majority of firms in the Canadian oil industry, defied numerous firms’ initial attempts at adoption and caused the substantial benefits that early adopters experienced to be deferred. Such research can highlight the effects of supporting resources, capabilities and processes that are often overlooked in favour of individual resources that can be uniquely linked to sustained advantages. For example, there is an extensive and growing list of studies focussing on capabilities that may be embedded in key individuals (e.g., Hitt, Bierman, Shimizu, & Kochhar, 2001; Hatch & Dyer,
Where these studies could be extended is to examine why firms with advantages from human capital are able to retain this valuable capital and also avoid having the rents created appropriated by this key stakeholder group (as appears to happen in many professional sports organizations). Resource-based theory poses a series of questions about such barriers to imitation or isolating mechanisms, which can contribute to both future theoretical perspectives as well as to improve managerial practice.

Beyond imitation, Barney (1991) notes that substitution is another mechanism by which competing firms may attempt to negate their resource disadvantages. This suggests that there may be a degree of equifinality where alternative resource bundles are able to generate similar levels of value creation and performance (Gresov & Drazin, 1997). Current research tends to ignore this possibility, pursuing analyses that examine the extent to which exact duplication has occurred to remove an advantage. While case examples of this phenomenon might exist, there is value in empirical RBV research considering this possibility when assessing the link between resources and value creation as well as generating a better understanding of why ex post limits to competition were overcome in this manner. Consideration of substitutes raises other research possibilities, such as whether duplication via substitutes tends to lead to the erosion of this advantage across all firms implementing the strategy, affects only the original sources of the advantage, or if substitutes tend also to have isolating mechanisms. Barney’s (1991) discussion of environmental shocks redefining the bases of advantage is open for similar analyses, with valuable resources under the original context potentially becoming weaknesses or rigidities in the newly defined industry structure.

CONCLUSION

The discussion presented here follows similar lines to those presented in our earlier work. After over a decade, there still seems to be considerable work to be done to establish resource-based theory empirically. Despite continuing critiques, empirical RBV research holds considerable promise for strategy and management as well as numerous other disciplines. However, in such research, there remains a need for data to be collected in a manner that is consistent with the RBV. Indeed, a key consideration to advance the RBV is the availability of data sufficiently detailed to enable researchers to establish, on the one hand, a clear link between resources and disaggregated performance attributable specifically to those resources, and
on the other hand, to make the connection between bundles of resources/capabilities and performance. Presently the lack of data, we think, is the biggest barrier for empirically establishing the RBV. At a minimum, the remedy entails careful sample selection and the collection of richly detailed longitudinal data, some of which may only be available by doing research in organizations and not just on organizations.

NOTE

1. For convenience, the term ‘resources’ is used throughout this discussion to refer to both resources and capabilities.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the research assistance provided by Louise Proctor in preparation for writing this manuscript.

REFERENCES


RESOURCE COMPLEMENTARITY, INSTITUTIONAL COMPATIBILITY, AND SOME ENSUING METHODOLOGICAL ISSUES IN TESTING THE RESOURCE-BASED VIEW

Tailan Chi and Edward Levitas

ABSTRACT

We argue that resource-based view (RBV) researchers must take into account three interdependencies, (i) intrafirm resource complementarity, (ii) interfirm resource complementarity or rivalry, and (iii) compatibility or incompatibility of firm resources to broader socio-economic institutions, when attempting to empirically verify the RBV. However, these interdependencies lead to three potential causes of statistical bias, which can reduce the interpretability of such empirical examinations. First, omitted variable bias results from a researcher’s inability to find and include in empirical analyses appropriate operationalizations of constructs. Second, selection bias can arise when a researcher samples only from one subset of the population, and not others. Bias in estimates can
occur if a correlation between unobserved determinants of the outcome and factors affecting the selection process exist. Finally, joint dependence, where two explanatory variables are themselves mutual determinants, can lead to biased estimation.

Without question, the resource-based view (RBV) of the firm is premised on the concept of value (Barney, 1991, 2002; Penrose, 1959; Priem & Butler, 2001; Rumelt, 1984; Wernerfelt, 1984). According to Barney (1991, p. 106), value stems from a resource’s ability “to exploit opportunities or neutralize threats in a firm’s environment (Barney, 1991, p. 106).” When the revenues from such exploitation exceed the firm’s opportunity costs associated with resource mobilization, the firm can appropriate or extract a portion of that value (Rumelt, 1987).

Some of the challenges in testing the RBV are widely recognized. As articulated by Reed and DeFillippi (1990), what sustains a firm’s superior performance is often a diverse assortment of resources (technology, marketing, manufacturing, etc.) that interact with one another in a complex manner to enhance profitability. Because the compositions of such resource bundles are firm specific and likely to vary across industries and over time, it is inevitably difficult for academic researchers to find empirical measures that achieve a good level of consistency and accuracy. Furthermore, another condition that is also considered necessary for any resource to sustain a firm’s superior performance is what Lippman and Rumelt (1982) call causal ambiguity – uncertainty about the causal connections between managerial actions and economic results. This condition makes it difficult for outsiders such as academic researchers to identify which resources constitute the relevant set for a given firm. In the face of these difficulties, empirical researchers generally have to rely on easily identifiable variables considered to reflect the strength of certain specific resources that are typically valuable for a firm to have, such as technology patents and research and development (R&D) and marketing expenditures. These “generic” measures are likely to capture only a subset of the resources that sustain a firm’s superior performance, thus subjecting the empirical model to errors in measurement due to likely the omission of measures for critical complementary resources from the model.

The purpose of this paper is to examine some methodological issues arising not only from complementarity among different types of firm resources but also from compatibility between a firm’s resources and the broader industry and societal institutional constraints a firm may face. We will scrutinize three types of interactions that should be considered in testing
the performance effects of firm resources. The first is intrafirm interactions among complementary resources that cannot be easily measured as a whole set, as explained in the previous paragraph. The second is interfirm interactions between potentially complementary resources that are in the possessions of different firms but can used to suppress the value of a focal firm’s resources if one of the other firms decides to utilize its resources as a competitive weapon. For instance, an existing consumer product giant like Sony or Matsushita can use its marketing or manufacturing prowess to limit or deter the adoption of a rivalrous technology from a startup company. The startup’s new technology can be at least as innovative and appealing as the giant’s own based on patents granted, expertise opinions and consumer surveys, but generate far less value due to rivalry from the giant. The third type of interaction we examine is the compatibility or lack thereof of a resource, or the output generated from the resource’s service (e.g., a new product), with existing socio-economic institutions. For instance, despite its innovativeness and high functionality, a new technology may be incompatible with existing industry standards, or its product (e.g., a family planning device) may find strong resistance from powerful social groups based on their religious and cultural norms.

Failure to recognize these interactions can lead to biased or inconsistent estimators when empirically testing the RBV. Specifically, when the researcher relies on easily identifiable variables to measure firm resources, other important variables that reflect other complementary resources and existing socio-economic institutions are likely left out of the empirical model. The omission of important explanatory variables can result in estimation errors being correlated with some of the explanatory variables of empirical models and, therefore, biased estimates. Problems associated with neglecting these interactions are not limited, however, to this “omitted variable” issue. Endogenous selection may also arise when, for example, a firm lacks critical complementary resources or when a new technology it attempts to develop is perceived to be incompatible with existing socio-economic institutions. In such instances, the firm may simply curtail its development effort in later stages of development. This can cause certain variables commonly used to measure firm resources (e.g., patents granted or R&D expenditures) to be partially determined by the possession of complementary resources or existing socio-economic institutions and thus become endogenous rather than exogenous variables. The endogeneity of these resource variables can again cause biases and inconsistencies in estimation.

More specifically, we pinpoint three causes of bias, which can plague empirical examinations of the RBV: omitted variable bias, selection bias,
and joint dependence. Omitted variable bias occurs when theory suggests the need for a researcher to include a control variable in a regression analysis, but she/he is unable to do so because of data unavailability. Selection bias, a form of omitted variable bias, may arise when a researcher samples only from one subset of the population, and not others. If unobserved factors affecting the response are correlated with unobserved factors affecting the selection process, empirical estimations may produce biased and inconsistent parameter estimates. Joint dependence refers to the instance where two variables are mutual determinants of each other. For instance, two sets of resources may affect levels of each other as well as a hypothesized outcome variable. R&D and patents may be mutually determining (R&D investments may lead to the production of patents, and patents may lead to R&D investments needed to move patented technology to a commercialization phase) and in combination, may promote firm performance. All three circumstances result in correlations between predictor variables and the error term, potentially leading to the production of biases and inconsistent parameter estimates.

Our paper proceeds as follows. First, we review the concept of value in the RBV framework and its relationship to the resource complementarity and institutional compatibility issues raised in the introduction. We then examine in greater detail how the intrafirm and interfirm resource interactions and compatibility or incompatibility with existing socio-economic institutions may lead to estimation biases and inconsistencies in the study of a resource’s performance effect. Third, we provide a more technical discussion of these methodological issues and their potential remedies. The last section summarizes and concludes the paper.

THE RBV AND DETERMINATION OF RESOURCE VALUE

The RBV enjoys mounting popularity in the management literature. Indeed, Barney’s (1991) work which identifies four attributes identified as necessary for a bundle of resources to sustain economic profit for the owner of the resources in question (value, rarity, inimitability, and non-substitutability) has achieved enormous attention in the management literature (Levitas & Ndofor, 2006). However, both theoretical and empirical issues remain with regard to the RBV (Priem & Butler, 2001; Levitas & Chi, 2002; Levitas & Ndofor, 2006).

When examining resource value, RBV theorists have generally defined it in terms of how resources affect firm success. Barney (1991) focuses on a
resource’s effect on promoting a firm’s ability to neutralize threats or exploiting opportunities. Makadok and Coff (2002) seem to equate resource value with the effect resources have on a firm’s ability to survive or thrive. These definitions imply two conditions that directly relate to subject of this paper. First, in order for the gain from deployment of a resource to exceed the cost of mobilization, the firm must hold the right to the quasi rent from the resource (Rumelt, 1987). If the resource in question resides in a specific functional area of the firm such as R&D or marketing, as is normally measured in empirical studies of the RBV, the resource must exhibit some complementarity with resources residing in other functional areas of the firm. Otherwise, the managers and employees in the specific functional area would be able to appropriate virtually all the quasi rent from the resource. Without any firm-specific factor anchoring these resources to the firm, the resources can be easily exported to another (competing) firm or otherwise eroded through wage bargaining. Second, the value of the resource is subject to broader industry and societal environments. In other words, resource utility cannot be estimated in isolation but must be assessed only within the societal, cultural, and network contexts in which that resource is (to be) employed. Consistently, after surveying much of the current RBV literature, Priem and Butler (2001, pp. 29–30; emphasis in the original) conclude that “it is the market environment, through opportunities and threats, that determines the degree of value held by each firm resource in the RBV. As the competitive environment changes, resource values may change.” Hence, extant theoretical formulations in the RBV clearly imply that resource efficacy is a function of the complementarity and institutional compatibility issues that we consider in this paper. However, we further argue that these complementarity and compatibility issues give rise to potential estimation biases and inconsistencies in testing the RBV.

In addition, resource development and deployment is an interactive dynamic process that fundamentally depends on what the firm does. In other words, resource accumulation and firm strategy are both endogenous to the firm’s activities. This is consistent with Schumpeter’s (1934, pp. 66–67) notion that value emerges from a process of creative destruction where the “carrying out of new combinations” of resources that result in the “competitive elimination of the old combinations”. Similarly, according to Penrose (1959), new utilizations and combinations of resources or their services obsolesce old ones in ways, which create societal benefit, and firms extract or appropriate value by exploiting new combinations in manners where the economic benefits exceed costs. From this perspective, if the variables that an empirical researcher uses to measure a firm’s
accumulated resources also reflect the strategic choices the firm makes (e.g., R&D expenditures may reflect the level of investment in uncovering new technologies as well as the decision to attempt such discovery), there would be an even greater chance for these variables to be dependent on other variables in the model that are intended to measure the firm’s resources or its competitive environment (e.g., lack of competing technological standards or government regulations that would forestall successful commercialization of the technology). This suggests another possible source of bias that should be taken into account in empirical modeling.

THREE TYPES OF INTERDEPENDENCIES

In this section, we discuss in some detail three types of interdependencies that variables intended to measure firm resources can have with variables that may be excluded from a model for testing the RBV. The interdependencies result from (i) intrafirm resource complementarity, (ii) interfirm resource complementarity or rivalry, and (iii) compatibility or incompatibility of firms resources to broader socio-economic institutions. In turn, these interdependencies can lead to the three causes of bias pinpointed in our introduction (omitted variable bias, selection bias, and joint dependence) which can plague empirical examinations of the RBV:

Intrafirm Resource Complementarity

Any resource employed by a firm, so long as it has an alternative use, engenders an opportunity cost. If a resource takes the form of knowledge and skills that are embodied in the employees of the firm, the opportunity cost is the sum of payoffs that the employees can gain from their alternative employment opportunities. In order for the resource to yield value for the firm, two conditions must exist. First, there is complementarity between the resource in question and some other resource with which the resource is being used jointly inside the firm. Second, the complementarity cannot be fully exploited by the individual possessor(s) of the knowledge and skills outside the firm. For instance, if a group of scientists working jointly in the firm’s research laboratories can create at least as much value by forming an independent research company themselves (and assuming no other contractual obligations tying them or the technology to the firm), then they could in theory extract all the value for themselves by bargaining with the
owners of the firm jointly. So, the joint of these two conditions implies either of the following:

1. The owners of the firm possess proprietary organizational skills or hold the rights to certain organizational routines that enhance the value created by the employees working jointly (Nelson & Winter, 1982).
2. There exist interunit complementarity across different functional areas of the firms, and hierarchical organization of the complementary activities within the firm achieves greater efficiency in exploiting the complementarity (Barzel, 1989; Williamson, 1985).

So, from this perspective, a likely perquisite for the firm to appropriate rents from its resources is the existence of complementarity among activities in different functional areas.

Furthermore, the presence of complex interactions among resources residing in different parts of the firm is also considered in the theoretical formulations of the RBV to be a main reason for a firm to sustain superior performance (Reed & DeFillippi, 1990). If a firm’s competitive advantage primarily resides in some narrow area of activities such as R&D or marketing, then a competitor can more easily replicate the advantage either through imitation or by hiring away the key employees conducting these activities. When there is strong complementarity among multiple areas of activities inside the firm, the price that a competitor must pay to bid away the key employees in a given area will be more prohibitive since the firm is able to share some of the value created with its employees.

This type of complementarity should pose no serious problem to an empirical researcher if she/he is able to identify the key components in a firm’s portfolio of complementary assets and is able to measure their strengths separately or obtain a reasonable composite measure for the overall strength of the portfolio. The basic theoretical proposition of the RBV, however, makes it inherently difficult for an empirical researcher to obtain precise and consistent measures for firm resources. As Lippman and Rumelt (1982) articulated, sustained superior performance by a firm must be characterized by what they call causal ambiguity – uncertainty about the causal connections between managerial actions and economic results. If it is easy for outsiders to identify what is exactly causing a firm’s superior performance, competitors would be able to imitate the firm’s advantage without too much difficulty, which in turn would erode the firm’s performance. In order for supernormal performance to persist, it must be difficult for outsiders to identify the sources of the firm’s competitive advantage. Such difficulty naturally also poses challenges for empirical researchers to obtain accurate and consistent
measures for firm resources (Levitas & Chi, 2002; Levitas & Ndofor, 2006). Consequently, many researchers rely on variables that are relatively easy to identify as reflecting the strengths of certain firm resources, such as patents granted and R&D and marketing expenditures. Unfortunately, such easily identifiable variables are unlikely to capture all the key components in the firm’s portfolio of valuable resources, and the omission of other important components because of measurement difficulty subjects the empirical model to potential biases in estimation.

Another potential problem from this type of measurement difficulty is that the variables that the researcher uses to measure firm resources may also represent the strategic choices made by the firm. Because resource development and deployment involves an interactive dynamic process (Schumpeter, 1937; Penrose, 1959), a firm’s resource accumulation and strategy choice are both endogenous to the firm’s activities. From this perspective, it is easy to conceive that some of the resource variables may represent a firm’s initial resource endowments and other variables also intended to measure firm resources may represent more of the firm’s strategic choices and are dependent on the initial endowments. To be more precise, let \( x \) denote a variable reflecting some initial endowment of the firm and \( z \) denote a variable representing a resource that the firm accumulated due to strategic choices made in the past. When both \( x \) and \( z \) are used to explain performance, we have

\[
\pi = a + \beta x + \gamma z + \epsilon
\]

and

\[
z = \eta + \kappa x + \nu,
\]

where \( \epsilon \) and \( \nu \) are residual errors from their respective regressions. This type of interdependency among the variables can make some of the intended resource variables endogenous as described here, and cause a more serious potential for biased estimation. An even more complex form of interdependency occurs when there is reason to believe that two (or more) resource variables represent complementary investments that tend to reinforce each other in a dynamic process of strategic choices. In such a case, there would be two (or more) resource variables in the model that are not only endogenous but also mutually interdependent, posing an even greater risk of biased estimation bias unless the simultaneity in the determination of the variables are properly modeled. We will provide a more technical discussion of the estimation biases associated with endogeneity in the next section.

**Interfirm Resource Complementarity or Rivalry**

The statistical problem can get even more complex if the value of a resource is highly dependent on its joint use with a complementary resource that is in the possession of another firm. In the absence of significant transaction costs, two
rational firms holding complementary assets should be able to find a way to exploit those resources easily through contracting, joint venturing, or merger or acquisition. High transaction costs, however, can make each of these possible arrangements for joint production infeasible. To illustrate this point, suppose that a startup company and an entrenched multinational corporation (MNC) developed competing technologies for a new product and that the startup’s technology is at least as good as or better than the MNC’s. Furthermore, suppose that the MNC possesses the requisite manufacturing and marketing capabilities to make and market the new product but the startup company does not. Under these conditions, the MNC would not be willing to pay much for the startup firm’s technology even if that technology is better than that of the MNC. The asymmetry in bargaining power can motivate the MNC to take a tough position aimed at extracting most of the rent from any collaborative deal with the startup. The owner of the startup could easily feel a sense of inequity and refuse to concede (Camerer & Thaler, 1995), causing an otherwise mutually beneficial deal between them to fall apart. As explained by Teece (1986), even if a deal is struck under such conditions, the innovator lacking the complementary resources can easily end up gaining little from its innovation. In addition, the entrenched MNC can even use its power over distributors to deter the adoption of the product made using its startup rival’s technology, further suppressing the value of the technology. It is worth noting here the presence of complementary resources across different firms in a rivalrous situation can turn potential complementarity into a negative influence on the value of a firm’s resources.

The type of interfirm resource complementarity and rivalry would not be much of a problem if the strengths of the various complementary resources were adequately measured in an empirical model. In the hypothetical case outlined in the preceding paragraph, the more valuable resource would simply be in the form of manufacturing and/or marketing prowess rather than the ability to generate technological innovations. However, if no measures for manufacturing and marketing resources were included in the empirical model due to measurement difficulties, then the performance effect of technological resources (e.g., measured as patents counts or R&D expenditures) would likely be either exaggerated or underestimated.

Moreover, the potentially competing firms can also be expected to alter their investments in a given resource (e.g., technology) depending on how strong their other complementary resources are. Specifically, a firm lacking in manufacturing or marketing resources is likely to foresee the difficulty in bargaining with a manufacturing or marketing giant and decide to invest less in technology development. This rational behavior then causes a firm’s
investment in one resource to be partially determined by its existing endowments in other resources. If the calculus of the firm is such that investment in one resource is not worth making unless its possession of some other complementary resource exceeds a certain magnitude, then we will witness no investment made in the resource from the firm that does not possess a sufficient magnitude of the other resource. This condition would give rise to selection bias in estimation that will be discussed in greater technical detail in the next section. Suffice it here to say that the presence of interfirm resource complementarity or rivalry many require more complex statistical models to be used in testing the RBV.

Institutional Compatibility

The creation of value from a resource entails its utilization in a production process, the transfer and sale of some output (product or service) thus generated to the consumer, and the actual consumption process. Both the gross benefits from consumption of the output generated from the resource and the costs associated with the production, transfer and consumption stages are dependent on the broad socio-economic institutions. The existing socio-economic institutions can be either favorable or unfavorable to the creation of value from the resource. Specifically, incompatibility with such institutions as industry standards, laws and regulations, and social and cultural norms diminish the value that can be created from the resource. The dependence of a resource’s value on the broad socio-economic institutions would not present any significant challenge to an empirical researcher if it varies in an essentially random manner across firms. There are, however, reasons to believe that the variation of firm resources in their compatibility with socio-economic institutions is not random.

Points exist in the life cycle of an industry where innovators simply do not possess the knowledge and/or have the environmental conditions necessary to create new technologies. One reason for such impingement is the establishment of “dominant designs” or industry standards that are adopted across product lines within an industry (e.g., the use of Microsoft operating systems in the majority of personal computers worldwide). Dominant designs can be so calcified as to preclude challenge from competing technologies (Tushman & Anderson, 1986). For several reasons, incumbent firms often have considerable interests in insuring their profitability and survival. In some cases, standards are based on technology owner’s own technologies. Owners may benefit from network externalities emanating
from technologies as the value of a product or service increases with volume of use. Accordingly, attempts to introduce competing standards will be met with considerable resistance by standard-owning firms (e.g., Salant, Gilbert, & Newberry, 1984).

Furthermore, establishment of and adherence to industry standards allows firms to fully exploit processes built around these technologies through experiential learning over time. The value of such experiential learning, however, may decrease if new standards are introduced as firms may have to investigate new technologies as well as new ways of successfully marketing that technology. Disruption of a standard, in other words, obsolesces investments made previously in learning.

The network of actors that has come to depend on a dominant design, however, is not limited to incumbent firms (Tushman & Rosenkopf, 1992). Other groups such as buyers can play an extremely important role in the establishment of standards. Making informed purchase decisions requires buyers to make considerable investments in understanding their own desires as well as understanding what will placate these desires. Search costs as well as learning investments (needed to understand and use the technology) are often required. Cumulatively, these costs may dissuade switching to products utilizing non-dominant technologies. Buyer power, therefore, can have a significant impact on technology introduction.

In the same manner, suppliers may also become averse to reorienting their systems and factories that adoption of new technologies is often precluded. Diffusion lags may preclude immediate acceptance and thus revenue streams when new superior technology is not consistent with currently acceptable technologies. Scarce or cost prohibitive capital needed to build necessary scale may also preclude the adoption of new technology, as the industry collectively may not be ready to welcome change until the full benefit from the new technology is widely understood.

In addition to industry standards, a change in government regulation can dramatically restrain or unleash competitive forces, and a lack of consistency with existing social norms can cause substantial resistance to the adoption of new technologies.

In the past decades, deregulations in industries such as commercial airline and banking have significantly increased both the potential rewards and risks facing industry participants, correspondingly widening the dispersion of firm performance. The greater dispersion of performance implies a strengthening of the effects that resources can have on firm performance. Conversely, government protection of certain industries or industry segments against foreign competition can have the opposite effects.
Although events such as changes in industry regulation happen infrequently, their influences on the performance effects of resources tend to be very substantial. A lack of control for these events in the empirical model can easily result in omitted variable bias.

In certain industries such as healthcare and agriculture, the sale and consumption of products and services are more subject to the influence of social norms than in other industries. Products that involve generically modified organisms or used in family planning, for instance, often encounter resistance from some social groups whose concepts of morality are in opposition to such technologies. Even though a new product or technology may be highly innovative and appealing based on technical data (e.g., patents generated, expert opinions, and consumer surveys), a lack of compatibility with the prevailing socio-cultural norms among even a relatively small segment of the population can make it difficult for the product or technology to yield value for the innovator. If a researcher focuses on an industry that is heavily influenced by prevailing social norms, a lack of control for a technology’s compatibility with such norms in variable selection can easily yield biased estimation. Since the perceived risk of encountering social resistance reduces the innovator’s expected reward, a resource variable that the researcher is able to measure may also be partially determined by the concern over the risk. The dependence of a firm’s investment in certain resources on compatibility with social and cultural norms can make the measured variable endogenous.

**METHODOLOGICAL PROBLEMS AND POSSIBLE REMEDIES**

In the previous section, we suggested that conditions of resource complementarity and institutional compatibility can easily give rise to the problems of omitted variable bias, selection bias, and joint dependence in empirical studies of the RBV. In this section, we examine these three methodological issues in greater technical detail and sketch a number of possible remedies.

**Omitted Variable Bias**

Suppose that the true underlying model linking firm activities to a performance outcome is in the following form:

\[ y = \beta_1 x_1 + \beta_2 x_2 + \epsilon \]  

(1)
If our analysis regresses $y$ on $x_1$ only leaving out $x_2$, then the estimator for $\beta_1$ would be

$$b_1 = \frac{\sum x_1 y}{\sum x_1^2}$$

$$= \beta_1 + \frac{\sum x_1 (\beta_2 x_2) + \sum x_1 \varepsilon}{\sum x_1^2}$$

(2)

after substituting the right-hand side of Eq. (1) for $y$. We have $E(b_1) = \beta_1 + \sum x_1 (\beta_2 x_2)/\sum x_1^2$ since $E(\sum x_1 \varepsilon) = 0$. Unless $\sum x_1 (\beta_2 x_2) = 0$, we have $E(b_1) \neq \beta_1$. So, the omission of a relevant variable causes biased estimation of the regression parameters for the included variables. In terms of testing the RBV, this means that the performance effects of the resource variables would be incorrectly estimated if some important components of the firm’s resource portfolio were not included in the model. Strict complementarity between two resource variables such as $x_1$ and $x_2$ means that the correctly specified model should have an interaction term $x_1 \cdot x_2$ in the regression equation, and leaving it out would also cause biased estimation.

To avoid such estimation biases, one should try to find measures for all the theoretically relevant components of a firm’s resource portfolio. Because of the inherent difficulty in finding reliable measures for firm resources, the downsides from including more variables is increased chance of measurement errors and reduced efficiency in estimation (Green, 2000). Although an efficient way to address measurement errors is the use of instrumental variables, this approach is likely difficult to apply given the inherent difficulty in finding variables that in theory correlate with various resource variables. The more applicable remedy, then, is to rely on a large sample size to keep the estimator consistent based on its asymptotic properties (Green, 2000, p. 120).

**Selection Bias**

In some industries (e.g., biotechnology and microelectronics), there are often a significant number of companies that do not yet have regular revenues even though they seem to possess strong resources in some areas (e.g., producing a large number of patents for their size). Without a regular revenue source, the accounting measures of their performance are necessarily poor, but the poor performance based on such measures may
be due to other factors than a weakness in the measured resource that also explain their lack of regular revenues. The omission of these factors (such as lack of certain complementary resources that are hard to measure) in the model could result in biased estimates.

This type of estimation bias can be corrected using a sample selection model of the genre that Shaver (1998) suggested for a study on the performance of foreign acquisition investments. To illustrate, suppose that the correct model is of the form shown in Eq. (1) above, but we do not have anyway to measure $x_2$ directly. Because we know that an observable attribute of the firm (e.g., no regular revenue stream) reflects $x_2$, we can use a selection equation with the observable attribute as the response variable to calculate a correction for the omission of $x_2$ from the model. Let $z$ be a binary variable that denotes the observable attribute, and regress it on a set of variables that are considered to affect the binary choice $z$.

$$z = \gamma_1 w_1 + \gamma_2 w_2 + u$$

The result from this regression produces a correction variable, commonly referred to as $\lambda$, which can be used to correct for the bias that would result from the omission of $x_2$ in the main model (Green, 2000, pp. 926–936).

A large variety of selection models have been developed in recent years to compensate for potential estimation biases resulting from many different types of deficiencies in the data, ranging from omission of important explanatory factors to incidental truncation of a variable. As discussed in the previous section, a firm’s investment in a particular resource may not occur if the strength of some other complementary resource in the firm’s possession falls below a certain threshold. In such a case, the variable intended to measure the firm’s investment in the resource is endogenous as well as truncated, and a sample section model can be used to correct the bias that arises from the endogeneity and truncation.

**Joint Dependence**

As noted earlier, variables that are used to indicate the strengths of some firm resources may reflect the accumulated investments in complementary resources over time due to the firm’s strategic choices. If this is the case, the variables become jointly dependent, and a model that does not take into account such joint endogeneity would yield biased estimates of their effects.
A simple illustration of such a relationship is as follows:

\[ y = \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon \]  
(4a)

\[ z_1 = \beta_1 x_1 + \gamma_2 z_2 + v_1 \]  
(4b)

\[ z_2 = \beta_2 x_2 + \gamma_1 z_1 + v_2 \]  
(4c)

where \( z_1 \) and \( z_2 \) can be viewed as two mutually reinforcing investments that affect profitability \( y \). In the equation system (4), both \( z_1 \) and \( z_2 \) are correlated with the error terms \( \varepsilon, v_1 \) and \( v_2 \), subjecting the estimates of their effects from ordinary least squares (OLS) regression to biases. The methods for dealing with such endogeneity are discussed in any econometric textbook and programmed into many popular statistical packages such as Stata, LIMDEP and LISREL. What we would like to point out here is that each model has a particular set of assumptions that may or may not be satisfied by the data that a given researcher has collected. So, it is important to select a model whose assumptions the available data do satisfy. In some cases, the researcher may need to revise the model using more basic programming commands in the statistical package in order to address peculiarities in the available data.

**SUMMARY AND CONCLUSION**

RBV researchers must take into account three crucial interdependencies when attempting to empirically verify the RBV. Accounting for these interdependencies, however, poses empirical difficulties for researchers. Intrafirm resource complementarity refers to the connections and mutually reinforcing dynamics among resources within the firm. Often what determines long-term uniqueness is not a single resource but the network of resource interactions internalized by the firm. Not only is such interconnectedness often necessary to extract value from a single asset, but also it promotes immobility of resources across firm boundaries (and thus uniqueness sustainability). Interfirm resource complementarity refers to extramural connections among resources. Often, a single firm does not have all complementary factors needed to extract value from its own resources. The firm, for example, may have technological prowess but the inability to commercialize its own technologies. In this sense, the value of its technological resources is contingent on the access to (distribution and marketing) resources of other (possibly competing) firms. The final
interaction we consider is the degree of compatibility between a firm’s resources and the broader socio-economic institutions in which the firm exists. Resource value is context dependent and is subject to the mores and whims of its external environment. Failure to account for legal statues, ethical debates and moral obligations, as such, can deplete technologically feasible resources of any economic value.

RBV researchers must account for these interdependencies. However, such accounting has the potential to cause statistical biases if not managed properly. Encountering such biases when empirically testing the RBV is highly likely given the RBV’s focus on causally ambiguous resources and resource interactions. First, omitted variable bias results from a researcher’s theoretical need to include a variable in a regression analysis, but her/his inability to do so. Such bias may occur, for example, due to a researcher’s inability to find a proper operationalization for manufacturing capabilities. Second, selection bias occurs when a researcher samples only from one subset of the population, and not others, and if unobserved factors affecting the response are correlated with unobserved factors affecting the selection process. This can occur, for example, in instances when researchers focus on pursued strategies but omit consideration of ignored strategies in their empirical analyses. Finally, joint dependence refers to the instance where two variables used to explain a third are mutual determinants of each other. Two resources that determine an outcome (e.g., performance) may themselves be mutually reinforcing. In all cases, such biases can reduce the interpretability of empirical results of the RBV.

In this paper, we have attempted to highlight key resource interdependencies, related empirical problems, and possible resolutions to these problems. The resolutions noted, however, are certainly not exhaustive, and empirical researchers should continuously search for techniques that can best address the realities of their research questions and their data. One should note that in addition to a limited number of solutions noted, some methods we highlighted may also be limited in certain circumstances such as in the presence of heteroscedasticity, with the utilization longitudinal data structures, or in instances where outcome variables on which the researcher focuses assumes certain non-normal distribution. Exhaustively covering all techniques is obviously beyond the scope of this paper (and perhaps most statistical texts). Nonetheless, our paper should serve as a reminder to researchers about factors to include in empirical analyses, as well as the possible difficulties when do so. In this sense, our goal is to alert researchers to these issues and problems in the hope that they will strive to incorporate them efficiently in their empirical analyses.
REFERENCES


MIXED METHODS IN STRATEGY RESEARCH: APPLICATIONS AND IMPLICATIONS IN THE RESOURCE-BASED VIEW

José F. Molina-Azorín

ABSTRACT

This chapter focuses on the empirical research on the resource-based view of the firm (RBV), and its main purpose is to analyse the use of mixed methods in this perspective. The recent advance of the RBV has posed new challenges, and the issue need not be quantitative versus qualitative methods, but rather how to combine the strengths of each in a mixed methods approach. This study carries out a literature review about the use of mixed methods in the RBV and provides an examination of opportunities and challenges associated with the application of mixed methods in order to improve RBV research. Moreover, the chapter seeks to introduce mixed methods research in order to familiarize strategic management and the RBV scholars about this type of research and its terminology, procedures, designs and purposes.
INTRODUCTION

The development of the strategic management field within the last decades has been dramatic. While its roots have been in a more applied area, the current field of strategic management is strongly theory based, with substantial empirical research (Hoskisson, Hitt, Wan, & Yiu, 1999). Strategic management relies on an array of complex methods drawn from various disciplines. The field is undergoing a rapid transformation in methodological rigor, and researchers face many new challenges about how to conduct their research and in understanding the implications that are associated with their research choices (Ketchen & Bergh, 2004). The choice between quantitative or qualitative methods has been the subject of controversy. Although strategy researchers employ both qualitative and quantitative approaches, the use of large sample, quantitatively-operationalised research designs dominates (Phelan, Ferreira, & Salvador, 2002; Rouse & Daellenbach, 1999). In any case, although the use of qualitative methods in strategy research has lagged significantly behind the use of more quantitative approaches, significant contributions to strategy theory and practice have come from qualitative studies (Barr, 2004). Moreover, with the increased popularity of qualitative research, an important methodological trend is to integrate qualitative and quantitative research methods (Hitt, Gimeno, & Hoskisson, 1998).

The acceptance and application of the resource-based view (RBV) of the firm (Barney, 1991; Grant, 1991; Peteraf, 1993; Wernerfelt, 1984) has been one of the most important developments in the field of strategic management (Hitt et al., 1998). The importance of the RBV may be due to two main reasons. First, this perspective helps to explain several corporate strategies such as diversification (Markides & Williamson, 1994; Tanriverdi & Venkatraman, 2005), international strategies (Carr, 1993; Hitt, Hoskisson, & Kim, 1997), vertical integration (Argyres, 1996; Leiblein & Miller, 2003) and alliances (Combs & Ketchen, 1999; Mowery, Oxley, & Silverman, 1996). Second, the RBV emphasizes the role that firm-specific resources play in creating competitive advantage and explaining heterogeneity in firm performance. Thus, the overall objective that informs the RBV is to account for the creation, maintenance and renewal of competitive advantage in terms of the characteristics and dynamics of the internal resources of firms (Foss, 1997). Building on the assumptions that strategic resources are heterogeneously distributed across firms and that these differences are stable over time, Barney (1991) argues that a firm has the potential to generate sustained competitive advantages and superior performance from
firm resources that are valuable, rare, inimitable and non-substitutable. While the value of the RBV as a theoretical framework is still being debated (Barney, 2001; Foss & Knudsen, 2003; Hoopes, Madsen, & Walker, 2003; Lavie, 2006; Makadok, 2001; Peteraf & Barney, 2003; Priem & Butler, 2001a, 2001b; Williamson, 1999), a growing number of empirical articles are appearing in the literature (Barney & Arikan, 2001; Nothnagel, 2005; Shimizu & Armstrong, 2004).

This chapter focuses on the empirical research on the RBV, and its main purpose is to analyse the use of mixed methods (Creswell, 2003; Tashakkori & Teddlie, 1998, 2003a) in this perspective. The recent advance of the RBV has posed new challenges, and the issue need not be quantitative versus qualitative methods, but rather how to combine the strengths of each in a mixed methods approach. The use of mixed methods can play an important role in the development of this perspective because results obtained from different methods have the potential to enrich our understanding of the problems and generate new insights regarding these issues.

The calls for the integration of quantitative and qualitative research methods have been carried out in the social sciences at large (Creswell, 2003; Greene, Caracelli, & Graham, 1989; Patton, 1990; Tashakkori & Teddlie, 1998, 2003a), in the management and organizational studies (Currall & Towler, 2003; Jick, 1979; Lee, 1991, 1999), in the strategic management field (Boyd, Gove, & Hitt, 2005; Hitt, Boyd, & Li, 2004; Hitt et al., 1998) and in the RBV (Henderson & Cockburn, 1994; Rouse & Daellenbach, 2002; Shimizu & Armstrong, 2004). For example, Shimizu and Armstrong (2004) believe that using both qualitative and quantitative methods best contributes to isolating potentially unobservable resources and testing RBV theory. To our knowledge, this is the first study to carry out a literature review about the presence of mixed methods in a specific strategic area. Two main contributions are made in this chapter. First, in the field of the RBV, this study analyses the use of mixed methods in the RBV and examines how mixed methods designs can contribute to the development of this strategic perspective. Second, in the field of mixed methods approach, the chapter proposes new ideas to advance understanding and knowledge in this area. Moreover, I seek to introduce this type of research to the uninitiated and, thus, to help strategy researchers become more familiar with mixed methods terminology, procedures, designs and purposes.

Different definitions have been provided about mixed methods research. Tashakkori and Teddlie (1998, pp. 17–18) point out that mixed methods studies are those that combine the qualitative and quantitative approaches into the research methodology of a single study or multiphased study.
Later, these authors indicate that mixed method research studies use qualitative and quantitative data collection and analysis techniques in either parallel or sequential phases (Teddlie & Tashakkori, 2003, p. 11). Johnson and Onwuegbuzie (2004, p. 17) indicate that mixed methods research is the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study. Tashakkori and Creswell (2007, p. 4), in the first issue of the Journal of Mixed Methods Research, define mixed methods as research in which the investigator collects and analyses data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or a program of inquiry. In this chapter, following Creswell, Clark, Gutmann, and Hanson (2003a, p. 212), a mixed methods study involves the collection or analysis of both quantitative and qualitative data in which the data are collected concurrently or sequentially, are given a priority, and involve the integration of the data at one or more stages in the process of research.

The rest of the chapter is organised as follows. Next, several aspects about mixed methods are highlighted to make clear what the literature in several and different social fields has said about this type of research. I then examine the RBV studies that have used a mixed methods design. The following section provides an examination of opportunities and challenges associated with the use of mixed methods that seeks to improve empirical research on the RBV. This section shows a comparison between RBV monomethod and mixed methods studies, suggestions about the use of mixed methods, and contributions of RBV mixed methods in the general field of mixed methods research. Finally, the last section offers a summary of the main conclusions.

MIXED METHODS RESEARCH

This section is going to deal with several issues related to the use of mixed methods research. Since this type of research has seldom been used in management and organizational studies (Currall & Towler, 2003; Lee, 1991), and therefore in the field of strategic management and in the RBV, an attempt is made to collect the main ideas provided by other social research fields such as evaluation, education, sociology, health sciences, nursing and psychology (Brewer & Hunter, 1989; Creswell, 1994, 2003; Creswell et al., 2003a; Creswell, Trout, & Barbuto, 2002; Forthofer, 2003; Greene & Caracelli, 1997; Greene et al., 1989; Hunter & Brewer, 2003; Morgan, 1998;
Morse, 1991; Newman & Benz, 1998; Onwuegbuzie & Leech, 2005; Patton, 1990; Rallis & Rossman, 2003; Reichardt & Rallis, 1994; Rocco et al., 2003; Tashakkori & Teddlie, 1998, 2003a; Twinn, 2003; Waszak & Sines, 2003). These works include key words, definitions, typologies and purposes of mixed methods designs. This facilitates the analysis of the mixed methods studies used in the RBV that will be offered in the next section and will later make it possible to identify the opportunities and challenges to improve empirical research and consequently the development of the RBV through a better, more widespread utilization of these methods.

During the last decades the methodological debate between quantitative and qualitative researchers changes its focus from the paradigm purity to the possibility of integration of qualitative and quantitative methods (Erzberger & Prein, 1997; Teddlie & Tashakkori, 2003). On the one hand, paradigm purists posited the incompatibility thesis with regard to research methods: compatibility between quantitative and qualitative methods is impossible due to the incompatibility of the paradigms underlying the methods. On the other hand, mixed methodologists posited the use of a different paradigm: pragmatism (Howe, 1988). A major tenet of Howe’s concept of pragmatism was that quantitative and qualitative methods are compatible. The concept of the triangulation of methods was the intellectual wedge that eventually broke the methodological hegemony of the monomethod purists (Tashakkori & Teddlie, 1998). Campbell and Fiske (1959) proposed the multitrait–multimethod matrix, which used more than one quantitative method to measure a psychological trait. Denzin (1978) also applied the term triangulation, and his concept involved combining data sources to study the same social phenomenon. He discussed four basic types of triangulation: data, investigator, theory and methodological triangulation. Jick (1979) discussed triangulation in terms of the weaknesses of one method being offset by the strengths of another. He also discussed “within methods” triangulation (such as multiple quantitative or multiple qualitative approaches) and “across methods” triangulation (involving both quantitative and qualitative approaches in the same study). Patton (1990) described three triangulation methods: reconciling qualitative and quantitative data, comparing multiple qualitative data sources, and multiple perspectives from multiple observers.

Tashakkori and Teddlie (2003) point out three areas in which mixed methods studies can be superior to monomethod approaches. First, mixed methods research can answer research questions that the other methodologies cannot. Although there is no necessary and perfect connection between purpose and approach, quantitative research has typically been
more directed at theory testing or verification, while qualitative research has typically been more concerned with theory building or generation (Punch, 1998). A major advantage of mixed methods research is that it enables the researcher to simultaneously answer confirmatory and exploratory questions, and therefore verify and generate theory in the same study. Second, mixed methods research provides better (stronger) inferences. Several authors have postulated that using mixed methods can offset the disadvantages that certain methods have by themselves. Johnson and Turner (2003) refer to this as the fundamental principle of mixed methods research: “methods should be mixed in a way that has complementary strengths and non-overlapping weaknesses”. Third, mixed methods provide the opportunity for presenting a greater diversity for divergent views. Divergent findings are valuable in that lead to a re-examination of the conceptual frameworks and the assumptions underlying each of the two (qualitative and quantitative) components.

From the original concept of triangulation emerged additional reasons for mixing different types of data (Creswell, 2003). Thus, regarding these reasons or purposes for combining qualitative and quantitative methods, Greene et al. (1989) reviewed 57 mixed methods studies in the field of evaluation and listed 5 purposes: (1) **Triangulation**, seeking convergence, corroboration, correspondence of results from the different methods. (2) **Complementarity**, seeking elaboration, enhancement, illustration, clarification of the results from one method with the results from the other method. The researcher uses different methods to assess different facets of a phenomenon, yielding an enriched, elaborated understanding of that phenomenon. (3) **Development**, the researcher uses the results from one method to help develop or inform the use of the other method. The salient feature of this design is the sequential timing of the implementation of the different methods. One method is implemented first, and the results are used to help select the sample, develop the measurement instrument, or inform the analysis for the other method. (4) **Initiation**, seeking the discovery of paradox and contradiction, new perspectives of frameworks, the recasting of questions or results from one method with questions or results from the other method. (5) **Expansion**, seeking to extend the breadth and range of inquiry by using different methods for different inquiry components.

Morgan (1998), in the field of health research, points out two main purposes: convergence (linked to triangulation) and complementarity. In his paper, this author focus on this last motivation for combining qualitative and quantitative methods, and he affirms that, unfortunately, the popularity of complementarity has been accompanied by a considerable amount of
confusion due to the lack of specificity in its definition. In his opinion, the key goal in studies that pursue complementarity is to use the strengths of one method to enhance the performance of the other method. Thus, Morgan’s concept of complementarity does not coincide with the definition proposed by Greene et al. (1989) (to analyse different facets of a phenomenon); it is actually more closely related to the concept of development by these authors.

The complexity of making design choices in mixed method studies reflects the confusion currently surrounding mixed method approaches (Tashakkori & Teddlie, 1998). Several authors have made attempts to create taxonomies of mixed method designs (Creswell, 2003; Greene et al., 1989; Morgan, 1998; Morse, 1991; Patton, 1990; Tashakkori & Teddlie, 2003b; Teddlie & Tashakkori, 2003). There are two main factors that help researchers to determine the type of mixed methods design for their study (Creswell, 2003; Morgan, 1998; Morse, 1991; Tashakkori & Teddlie, 1998).

- Implementation of data collection. Implementation refers to the sequence the researcher uses to collect both quantitative and qualitative data. The options consist of gathering the information at the same time (concurrent, simultaneous or parallel design) or introducing the information in phases (sequential or two-phases design). In concurrently gathering both forms of data, the researcher seeks to compare them to search for congruent findings. When the data are introduced in phases, either the qualitative or the quantitative approach may be gathered first, but the sequence relates to the objectives being sought by the researcher. Thus, when qualitative data collection precedes quantitative data collection, the intent is to first explore the problem under study and then follow up on this exploration with quantitative data that are amenable to studying a large sample so that results might be inferred to a population. Alternatively, when quantitative data precede qualitative data, the intent can be to test variables with a large sample and then to explore in more depth with a few cases during the qualitative phase.
- Priority (status/importance/dominance). The mixed methods researcher can give equal priority to both quantitative and qualitative research, emphasize qualitative more, or emphasize quantitative more. This emphasis may result from research questions, practical constraints for data collection, the need to understand one form of data before proceeding to the next or the audience preference. Mixed methods designs can therefore be divided into equivalent status designs (the researcher conducts the study using both the quantitative and the qualitative approaches about
equally to understand the phenomenon under study) and dominant-less dominant studies or nested designs (the researcher conducts the study within a single dominant paradigm with a small component of the overall study drawn from an alternative design). Sometimes, researchers and readers make an interpretation of what constitutes priority, a judgement that may differ from one inquirer to another.

These two dimensions and their possible combinations can lead to the establishment of several designs, which are represented using the notation proposed by Morse (1991, 2003). In her system, the main or dominant method appears in capital letters (QUAN, QUAL) whereas the complementary method is in lowercase letters (quan, qual). The notation “+” is used to indicate a simultaneous design, and the arrow “→” stands for sequential design. Thus, the following mixed methods designs can exist using these two dimensions (Johnson & Onwuegbuzie, 2004):

- Equivalent status/sequential designs: QUAL → QUAN; QUAN → QUAL.
- Equivalent status/simultaneous design: QUAL + QUAN.
- Dominant/sequential designs: qual → QUAN; QUAL → quan; quan → QUAL; QUAN → qual.
- Dominant/simultaneous designs: QUAL + quan; QUAN + qual.

**MIXED METHODS RESEARCH AND THE RBV**

The use of qualitative methods in the RBV empirical research is rare (Nothnagel, 2005) and, therefore, the expectations are that mixed methods will be used even more rarely. This is why a broad definition of mixed methods research has been used and an effort has been made to carry out a thorough search for relevant published studies. However, finding them requires some creative searching through the literature because the actual terms used to denote a mixed methods study vary considerably (triangulation; mixed methods; multimethods; qualitative and quantitative) (Creswell et al., 2003a).

A review of the literature about mixed methods in the RBV published between 1984 (the expression “resource-based view” originated in an article of Wernerfelt, 1984) and 2006 has been carried out. The search and collection of the studies was initiated by computerised searches of two databases, ABI/INFORM Global and Science Direct. The keywords chosen were “resource”, “capabilities”, “competence”, “triangulation”, “mixed method”, “multimethod”, “quantitative” and “qualitative”, used in
adequate combinations. Moreover, a manual search for relevant articles in all the issues of *Strategic Management Journal* between 1984 and 2006 followed. Finally, I examined the reference sections of existing reviews about RBV to manually search for possible studies. Many of the studies found, especially when the term “triangulation” was used, applied several methods but within the same approach (several quantitative data collection techniques or several qualitative data collection techniques). These studies were consequently ruled out for not being real mixed methods (quantitative and qualitative methods).

One way to organize and describe the mixed methods research used in the RBV is to examine the research design and the purposes of these studies. Table 1 shows the classification of our sampled mixed methods studies, which are grouped according to the designs mentioned in the previous section. A variety of purposes and designs have been used in the empirical research on the RBV with mixed methods. The authors of the original studies did not use these design names or types. In fact, it must be pointed that very few of these works are explicitly described as mixed methods studies by their authors. My designation of these studies as particular types is based on an ex post analysis of their characteristics. Deciding whether one of the methods is the dominant or whether both of them have an equal status turns out to be difficult sometimes. As can be seen in Table 1, not all the possible designs have been used. Moreover, the qual→QUAN design has been the most often used mixed methods research.

One approach to learning about mixed methods research designs in a particular field is to begin with the analysis of several mixed methods studies and explore the features that characterize them as mixed methods research (Creswell et al., 2003a). Therefore, I identify and review one exemplary study in each type of design.

As an illustrative example of QUAL→QUAN design (sequential, equivalent status), Sharma and Vredenburg (1998) carried out a study conducted in two phases within a single industry context (oil and gas industry). The first phase (exploratory) involved comparative case studies through in-depth interviews in seven firms in the Canadian oil and gas industry to ground the RBV of the firm within the domain of corporate environmental responsiveness. Interview data were triangulated through a qualitative content analysis of corporate public documents such as annual reports, environmental reports, company newsletters and newspapers reports. This exploratory study was intended to examine linkages between environmental strategies and the development of capabilities, and understand the nature of any emergent capabilities and their competitive
### Table 1. Mixed Methods Studies in the RBV.

<table>
<thead>
<tr>
<th>QUAL → QUAN</th>
<th>QUAN → QUAL</th>
<th>QUAL + QUAN</th>
<th>qual → QUAN</th>
<th>quan → QUAL</th>
<th>QUAL + quan</th>
<th>QUAN + qual</th>
</tr>
</thead>
</table>

**Note:** “QUAL” or “qual” stands for qualitative, “QUAN” or “quan” stands for quantitative; “→” stands for sequential design; “+” stands for simultaneous design; capital letters denote high priority or weight, and lower case letters denote lower priority or weight.
outcomes. This first phase ends with two hypotheses based on previous literature and this qualitative study. The second phase (confirmatory) involved testing the emergent linkages through a mail survey-based study of the Canadian oil and gas industry. The final written report is structured in two main parts: the exploratory study includes several sections (qualitative data collection, qualitative data analysis and results with the proposed hypotheses) and then the confirmatory study is carried out (with a quantitative data collection, quantitative analysis section and the results). The design of this mixed methods study is two-phase sequential study (first qualitative and second quantitative), with equivalent status design in our opinion. Qualitative phase helps to know the industry, and develop theory, hypotheses and the measurement instrument.

Dyer and Hatch (2006) is an example of the QUAN→QUAL design. Their study examined the role of network knowledge resources in influencing firm performance, using a sample of U.S. automotive suppliers selling to both Toyota and U.S. automakers. The findings empirically demonstrated that network resources had a significant influence on firm performance. Moreover, some firm resources and capabilities were relation-specific and were not easily transferable to other buyers or networks. From a point of view of mixed methods research, the authors indicated the dual nature of the research investigation. The first objective was to empirically examine the relationship between customer-to-supplier knowledge-sharing activities and the rate of improvement in supplier network performance. Thus, this quantitative part tested the hypothesis that a buyer that provides greater knowledge transfers to its supplier network will develop the suppliers’ production capabilities such that the suppliers’ operations for that particular buyer will be more productive. A survey was sent to the plant managers at the 97 suppliers in the Toyota’s U.S. supplier association called Bluegrass Automotive Manufacturers Association. The empirical findings from this part of the investigation confirmed that Toyota’s supplier network does produce components of higher quality and at lower cost for Toyota that for their largest U.S. customers. Then, the second objective (qualitative) was to explore why the supplier performs better as a member of one network (i.e., Toyota’s) than another network (i.e., GM, Ford, or Chrysler). Thus, in this part the authors examined why the network resources created by Toyota do not flow to GM even though GM has ties with that same network. Interviews were done at 13 suppliers to explore quantitative results and analyse this second objective. Therefore, in this study the authors used different methods to assess different facets of a phenomenon, yielding an enriched, elaborated understanding of that phenomenon.
Tripsas (1997) used QUAL + QUAN design, and this combination of quantitative and qualitative analysis was carried out to unravel the process of creative destruction. She argued that the ultimate commercial performance of incumbents vs. new entrants is driven by the balance and interaction of three factors: investment, technical capabilities and appropriability through specialised complementary assets. These issues were examined through a study of the technological and competitive history of the typesetter industry for a period of over 100 years. The core of the data consists of a comprehensive longitudinal data set covering the entire history of the worldwide typesetter industry from the inception of the industry in 1886 through 1990. It was collected during a 14-month field-based study. The data set included the entry date of every firm in the industry and, for those firms that exited, the exit date. Detailed data for 95 per cent of the products introduced by these firms covers product performance characteristics, price and unit sales over time. Firm-level data were supplemented by aggregate industry-level data, including industry size and growth. This quantitative data were supplemented with qualitative data about how organizations responded to new technology, including in-depth case studies of multiple firms. Detailed schematics of typesetter machines from each generation of technology were reviewed with development engineers in order to understand changes in machine components and architecture. This work enabled a careful determination of the nature of technological change and its effects on organizational competence. These data come from a combination of primary and secondary sources including company and trade association archives, field interviews, personal records of retired employees, industry consultants, industry historians, government records as well as industry trade and scientific journals. Both qualitative and quantitative analysis confirmed that established firms were handicapped by their prior experience in that their approach to new product development was shaped by that experience. Thus, triangulation is achieved in this study.

Yeoh and Roth (1999) is an illustrative example of qual → QUAN design, and they tested a model of the relationships among firm resources, firm capabilities and sustained competitive advantage between 1971 and 1989 in 20 firms in the U.S. pharmaceutical industry. They carried out a quantitative analysis, though they explained that the variables and their measurements used in the quantitative analysis were determined through a qualitative two-stage process. First, the secondary literature was reviewed to determine the resources and capabilities that are important for competition in this industry. These secondary sources included company annual reports, books on the industry, articles in the marketing, management and innovation
journals about this industry and medical journal articles. The second stage involved interviews with product and marketing managers at several pharmaceutical companies as well as with industry experts at the Pharmaceutical Manufacturers’ Association. In these interviews, informants were requested to identify the types of capabilities that they felt were critical for future success in the industry. Thus, this qualitative phase helps to carry out the quantitative study, which is the dominant part of the article. The research question is quantitative. Additionally, and contrary to what was done by Sharma and Vredenburg (1998), the qualitative part does not appear in such a deep, systematic and methodical way, without constructing any sets of in-depth cases. In any case, the qualitative phase also helps to know the industry and facilitate the quantitative part.

Although it is not an empirical work, the study of Rouse and Daellenbach (1999) has been included in Table 1. These authors provided ideas to develop a new mixed methods design: quan→QUAL. As said above, they emphasised the importance of qualitative research to stimulate and guide future studies of resource-based competitive advantage. Their framework begins with a quantitative four-step firm selection process: (1) selecting of a single industry; (2) clustering firms by strategic type or group within this industry; (3) comparing performance indices within strategic groups; and (4) identifying those firms within each strategic group that are the high and low performers. Then, these firms would be selected as research subjects using in-depth fieldwork or ethnographic study methods, that is, a qualitative inquiry. Given the contention that sustainable competitive advantages are organizational in origin, tacit, highly inimitable, socially complex, embedded in process, and often driven by culture, there can be no other way to obtain the data of interest that using a qualitative orientation. Fieldwork, which takes the researcher into the organization is essential to gain an in-depth knowledge and understanding of the organization and its processes.

As an exemplar of QUAL+quan design, Yauch and Steudel (2003) analysed an important intangible resource, specifically the organizational cultures of two small manufacturers using qualitative and quantitative data. The article described not only how qualitative and quantitative data contributed to the validity of the results through triangulation but also how the qualitative and quantitative research paradigms were used in a complementary fashion to produce a more complete understanding of the organizational cultures. Using methods from both research paradigms enabled a greater understanding of cultural artefacts and behaviours but more important of the underlying cultural values and assumptions. A prospective exploratory case study approach was used to examine the
impact of organizational culture on the cellular manufacturing conversion process. The ultimate goal of the research was to identify key cultural factors that had a positive or negative impact on the process of converting from a traditional functional manufacturing system to cellular manufacturing. Cultural assessment of the organizational values, assumptions and behavioural norms was accomplished through qualitative and quantitative means. Qualitative assessment of culture was accomplished through document review, participant observation and group interviews. The Organizational Culture Inventory, a cultural assessment survey, was used as an additional measure of organizational culture at each company. This study relied most heavily on qualitative data but supplemented it with quantitative survey results. In any case, the authors point out that if considered from a paradigm perspective, the research was primarily quantitative because the ultimate goal was to identify factors that could aid or hinder the implementation of cellular manufacturing for small manufacturers, a causal explanation generalised to a broader population. For this reasons, this study appears in two columns in Table 1.

Finally, Hoetker (2005) can be considered an exemplar of the QUAN+qual design. This author studied how firms select a supplier for an innovative component, using three literatures: transaction cost economics, inter-firm relationships and firm capabilities. The primary data source was the COMTRAK database that allowed to compile a complete inventory of quantitative data. The author supplemented this quantitative data with interviews at three notebook computer manufacturers in the United States and Japan. At the two firms that also produce displays, Hoetker interviewed individuals in both the notebook computer and display divisions. The author also interviewed principals at three major consulting firms in the industry, relevant officials at Japan’s Ministry of International Trade and Industry and other long-time industry participants. The combination of qualitative and quantitative methods allows to capture a more complete, holistic and contextual portrayal of the phenomenon.

**IMPLICATIONS FOR FUTURE RESEARCH**

*Mixed Methods Research vs. Monomethod Research in the RBV*

Quantitative and qualitative methods have been used in the empirical research on the RBV (Barney & Arikan, 2001; Hitt et al., 1998; Hoskisson et al., 1999; Nothnagel, 2005; Shimizu & Armstrong, 2004). In general, as
said above, although there is no perfect connection between purpose and approach, quantitative research has typically been more directed to theory testing, while qualitative research has typically been more concerned with theory building. Thus, most quantitative research is confirmatory while much qualitative research is exploratory. Creswell et al. (2003a) point out that in quantitative research, investigators ask questions that try to confirm hypotheses, with a focus on assessing the relationship or association among variables with a large sample. These hypotheses are assessed using instruments, observations or documents that yield numerical data. These data are, in turn, analysed descriptively or inferentially so as to generate interpretations that are generalisable to a population. In qualitative research, the inquiry is more exploratory, with a strong emphasis on description and with a thematic focus on understanding a central phenomenon with a few cases. Open-ended data collection helps to address questions of this kind through procedures such as interviews, observations and documents. Researchers analyse these databases for a rich description of the phenomenon as well as for themes to develop a detailed rendering of the complexity of the phenomenon, leading to new questions and personal interpretations made by the inquirers. In sum, quantitative methods offer a broad, generalisable set of findings presented succinctly and parsimoniously. By contrast, qualitative methods typically produce a wealth of detailed information about a much smaller number of cases.

The same as in the field of strategic management, quantitative studies prevail over qualitative ones in empirical research on the RBV (Nothnagel, 2005). The predominance of more quantitative-based methodological tools in the development of strategic management and the RBV does not mean that this approach is applicable to all research questions. The bulk of empirical resource-based work has focused on identifying resources that have the attributes that RBV predicts will be important for competitive advantage and firm performance, and then examining whether or not the predicted performance effects exist (Barney & Arikan, 2001). The performance effects of a wide variety of different types of firm resources have been examined, including human resources (Huselid, Jackson, & Schuler, 1997), technological resources (Helfat, 1997) or reputation (Deephouse, 2000). These studies use a quantitative approach to analyse the linkage between resources and results obtained.

However, certain questions and problems are more consonant with qualitative methods. Patton (1990) argues that qualitative research is adequate whether it is focused on process (how something happens) rather than on the outcomes or results obtained. For example, it is very interesting in the RBV
to know how resources, capabilities or competences emerge and develop inside the firm, examining the process of evolution (Ethiraj, Kale, Krishnan, & Singh, 2005; Hoopes et al., 2003). Hoskisson et al. (1999) point out that the case study methodology may be appropriate for the RBV research because it can provide much richer information about the firms’ idiosyncrasies. More sophisticated case methodologies have been adopted by Carr (1993), Collis (1991), Danneels (2002), Doz (1996), Helfat and Raubitschek (2000), Kotha (1995), Langlois and Steinmueller (2000), and Zander and Zander (2005), including detailed field-based case studies with collection of both archival and interview data. Moreover, Godfrey and Hill (1995) indicate that qualitative methodologies such as multiple case studies, event histories and ethnographic inquiries, may represent the best way forward in observing the effects of otherwise unobservable, idiosyncratic effects on business strategy and performance, such as those predicted by the RBV.

In any case, both quantitative and qualitative research methods have supporters and critics. For instance, Rouse and Daellenbach (1999) are critical of approaches to RBV research using quantitative analysis on large sample observations with secondary data, as those approaches are unlikely to be able to disentangle the variety of effects and fail to isolate sustained source of advantage. They recommend using a detailed, field-based comparison of carefully selected firms in an industry to uncover sources of competitive advantages. However, Levitas and Chi (2002) suspect that this qualitative approach has its own limitations if the value of observable variables is ignored and efforts to verify conclusions are denied. Thus, weaknesses can be found in both quantitative and qualitative research (Yauch & Steudel, 2003). Therefore, a first option or way to improve empirical research on the RBV consists in trying to overcome the possible limitations and weaknesses of monomethod quantitative and qualitative studies. On the one hand, as Levitas and Chi (2002) put it, in quantitative studies researchers need not to de-emphasize large sample methods, but rather to strive for creativity in operationalising constructs. On the other hand, those researchers who carry out qualitative studies must overcome some of the limitations attributed to this type of studies, such as subjectivity, personal biases and anecdotalism (Bryman & Bell, 2003; Shimizu & Armstrong, 2004; Silverman, 2002), trying to carry out data collection and analysis methodically and increase the quality of this type of research.

A second and related way to improve empirical research on the RBV is the use of mixed methods. Mixed methods research can answer research questions that the other methodologies cannot (Teddlie & Tashakkori, 2003). Thus, a single study can simultaneously examine a quantitative and a
qualitative research questions, and the researcher can assess and analyse different but related facets of a phenomenon, yielding an enriched understanding of it. For example, Dyer and Hatch (2006), as an example of the QUAN→QUAL design, examined the role of network knowledge resources in influencing firm performance, and their investigation had two objectives: (1) a quantitative objective was to empirically examine the relationship between customer-to-supplier knowledge-sharing activities and the rate of improvement in supplier network performance; (2) a qualitative objective was to explore why the supplier performs better as a member of one network than another network. Moreover, a major advantage of mixed methods research is that it enables the researcher to simultaneously answer exploratory and confirmatory questions, and therefore generate and verify theory in the same study. In our review of RBV studies, Sharma and Vredenburg (1998) carried out an exploratory study in a first phase, that was intended to examine linkages between environmental strategies and the development of capabilities, and understand the nature of any emergent capabilities and their competitive outcomes. This first phase ended with two hypotheses. The second phase tested the emergent linkages through a survey study.

On the other hand, mixed methods designs are also useful if they provide opportunities to better answer our research questions, even if those questions are linked to a pure approach. This idea can be applied to the RBV. Thus, taking into account that resources and competitive advantages are context-specific (Ethiraj et al., 2005), the task of finding a better answer to quantitative research questions could be made easier if, prior to the quantitative inquiry, a qualitative phase were carried out with the aim of acquiring a deeper understanding of that industry context. This would make possible a better knowledge of the strategic resources in that industry as well as the specific dependent variables, permit the design of a better measuring instrument, and facilitate and enhance the interpretation of the results obtained. The QUAL→QUAN and qual→QUAN studies try to achieve that. In most studies, there are no exploratory questions; simply, a confirmatory/quantitative question is answered more accurately when a previous qualitative phase has been carried out than when that is not the case.

This can be due to some inherent problems in quantitative RBV research. Thus, the RBV has been difficult to test empirically because many of the intangible constructs it purports to represent are difficult to operationalize (Barney, Wright, & Ketchen, 2001; Godfrey & Hill, 1995). This difficulty increases when an attempt is made to assess the attributes that resources
must have if they are to create and sustain competitive advantage; namely, being valuable, rare, inimitable and non-substitutable. Moreover, most of the research has used coarse-grained measures of firm resources (Hitt et al., 1998). Some researchers have used such variables as research and development intensity, advertising intensity and patents to proxy intangible resources (Almeida, 1996; Bierly & Chakrabarti, 1996; Chatterjee & Wernerfelt, 1991; Lee, Lee, & Pennings, 2001; Mowery et al., 1996). Other works use managers’ perceptions (King & Zeithaml, 2001; McEvily & Zaheer, 1999; Spanos & Lioukas, 2001; Zander & Kogut, 1995). Three additional issues have been problematic in the RBV quantitative research: the level of analysis, the dependent variable and the need for longitudinal studies (Rowe, Rouse, & Riaz, 2005).

Mixed methods research with a sequential design where the qualitative phase is carried out before the quantitative one, may solve these issues. A starting point is the number of industries included in a study. Some works analyse firms belonging to several industries (Farjoun, 1994; Markides & Williamson, 1994; Robins & Wiersema, 1995) whereas others focus on a single industry (Brush & Artz, 1999; Douglas & Ryman, 2003; Henderson & Cockburn, 1994; Miller & Shamsie, 1996; Pisano, 1994; Yeoh & Roth, 1999). Priem and Butler (2001a) argue that little work has been done with respect to evaluating the RBV in appropriate contexts. Thus, when the RBV is used as a framework to analyse the linkage between firm resources and competitive advantage and/or firm performance, researchers should focus on an industry. Examining a single industry arguably reduce the generalizability of the results but support more accurate measurement of firm-specific resources and their impact on specific firm performance adequate for the industry analysed (Hatch & Dyer, 2004; Hitt et al., 1998; Macher & Mowery, 2001). In fact, the qualitative part may play an important role for determining these independent (resources, capabilities, competences) and dependent (competitive advantage, performance) variables. Ethiraj et al. (2005) indicated that firm capabilities are context-specific and fruitful research in this area might emanate from enjoining an in-depth study of the capabilities specific to a context and careful empirical estimation of their significance and value.

For example, King and Zeithaml (2001), in their qual→QUAN design, pointed out that on-site interviews were held with the chief executive to generate a comprehensive list of specific competencies. Yeoh and Roth (1999) indicated that the identification and measurement of valuable resources and capabilities is difficult because of the industry specificity of these assets, and they indicate that the variables and their measurements
were determined through a two-stage process. First, the secondary literature was reviewed to determine the resources and capabilities that are important for competition in the pharmaceutical industry, and the second stage involved interviews with managers and experts, that were requested to identify capabilities that they felt were critical for future success in the industry. Douglas and Ryman (2003) utilised highly qualified industry experts to evaluate the hospital resource endowments with regard to their relative strategic value. Ray, Barney, and Muhanna (2004) pointed out that interviews with managers in insurance companies suggested several resources and capabilities that can influence the performance of the customer service process in an insurance firm.

Another important issue in the field of the RBV is the level of analysis (Peteraf & Barney, 2003). Few studies have been able to isolate potential sources of competitive advantage at the resource level. Since the RBV predicts that resources within business units generate value, it is important to examine the value generated by specific, critical resources, yet these effects have not been analysed directly. Precisely because the RBV asserts that the resources of interest are characterised as valuable, rare and inimitable, research methods need to be able to tap into the resource level for analysis (Rowe et al., 2005). Barney and Mackey (2005) point out that the most correct level of analysis at which to examine the relationship between a firm’s resources and its strategies is at level of the resource, not the level of the firm. Some mixed methods studies take into account this aspect (Ethiraj et al., 2005; Henderson & Cockburn, 1994; Ray et al., 2004). In a qual→QUAN design, the qualitative phase can help to determine this level of analysis. For example, Ethiraj et al. (2005) conceptualized the notion of capabilities at a more micro-level within the firm, namely the projects which the firm executes for its clients, developing appropriate measures for these capabilities, and examining their evolution and impact on performance. These authors pointed out that it appears that measuring capabilities at micro-levels within a firm is quite promising because it not only helps better estimate their economic significance but also provides clear guidelines to the firm on where and how it needs to improve its capabilities. Measures of capabilities at the aggregate firm level, while useful in identifying between-firm differences, provide little understanding of the micro-foundations of such differences. However, measurement of firm capabilities at the micro-level holds promise of enhancing our understanding about how and why some firms perform better than others.

Closely linked to the level of analysis is the consideration of the dependent variable (Rowe et al., 2005; Shimizu & Armstrong, 2004). Using a highly
aggregated dependent variable such as firm performance may obscure or misrepresent the contribution of a particular resource to competitive advantage. For example, one resource that has the potential to generate competitive advantage may be offset by another that has negative effects. Similarly, one resource that has the potential to generate value may be neutralized by dysfunctional organizational processes. Therefore, it is more valid to operationalize the dependent variable at the same level as the resource of interest and, further, to disaggregate that dependent variable so that the potential impact of other resources are controlled for (Barney, 2001; Priem & Butler, 2001a). Mixed methods research can facilitate the use of adequate dependent variables. Thus, a first qualitative phase can help to identify and measure these dependent variables, and then they can be used in the quantitative part. Several RBV qual→QUAN studies identified in our review take into account this issue (Ethiraj et al., 2005; Henderson & Cockburn, 1994; Ray et al., 2004).

Another important issue in the RBV empirical research is the need for longitudinal studies (Rowe et al., 2005). Barney (2001) points out that dynamic research is particularly important. Time is not only a factor in the path-dependent generation or development of resources and capabilities (Dierickx & Cool, 1989; Hoopes et al., 2003), but there may be a time-lag between resource utilization and performance (Priem & Butler, 2001a). Mixed methods can also help to face this aspect. Ethiraj et al. (2005) found evidence that an improvement in the productive deployment of resources over time yielded increases in project profitability.

On the other hand, in qual→QUAN and QUAL→QUAN designs, the qualitative data of the first phase can be used to enhance the interpretation of quantitative results. Therefore, another additional qualitative phase is not used (and thus the design is not qual→QUAN→qual or QUAL→QUAN→qual); actually the information collected in the first qualitative phase is used to help understand quantitative results. For example, King and Zeithaml (2001), in the discussion of quantitative results, refer to previous interviews with managers to comment these results.

Finally, Yeoh and Roth (1999) indicate that resource-based arguments recognize that different types of tangible and intangible assets are the source of firm advantage, and research often will identify the specific assets or stocks critical within a particular industry context. However, within a set of identified resources, additional understanding is needed regarding the different roles of the resources. These authors found that one set of resources is valuable because it contributes to transforming another set of resources. This finding suggests that the RBV may need to recognize the
hierarchical structure of resources as they relate to competitive advantage. A combination of qualitative and quantitative research may help to discover this hierarchical structure. Thus, Yeoh and Roth (1999) used a combination of interviews with managers and experts, and a structural equation model to analyze this issue.

In sum, in the RBV studies, the use of a qualitative part before the quantitative one may permit develop or extend theory that can be tested with the quantitative approach, identify the industry-specific resources and competences as well as the dependent variables, improve the measurement instrument of the quantitative phase (e.g., a questionnaire), determine the adequate level of analysis, facilitate the development of longitudinal studies, enhance the interpretation and understanding of quantitative results, and recognize the hierarchical structure of resources and capabilities.

Future Challenges of Mixed Methods Research in the RBV

Strategic management, and thus the RBV, is generally acknowledged to be one of the younger subdisciplines within the broader management domain (Boyd et al., 2005). Qualitative research complements quantitative research, and in tandem, quality research of both types can move the field forward more rapidly. The use of this mixed methods research in the RBV is incipient. When evaluating the findings of this review, the reader should keep in mind the broad definition of mixed methods that has been used. Moreover, the expression “mixed methods” only was found in two studies (Wei & Morgan, 2004; Yauch & Steudel, 2003), although expressions about the integration or combination of qualitative and quantitative approaches was showed in more studies. In any case, we have to take into account that other social sciences have more tradition in using this type of research. The key point in this section is to determine what we can learn of these disciplines. Therefore, my purpose is not to criticize previous work but to identify needs for future mixed methods research designs.

It was observed in the literature review section that the mixed methods studies performed in the field of the RBV had not used all the possible designs established in the section devoted to mixed methods research. Thus, the potential application of these and other designs used in other fields represents an opportunity for the development of the RBV. For instance, Currall, Hammer, Baggett, and Doniger (1999) used a QUAL→QUAN→QUAL design. Thus, their study’s use of qualitative and quantitative information in three phases promoted both “discovery” (i.e., theory
development) and “justification” (i.e., theory evaluation) and facilitated a “discovery-justification-discovery cycle” that was particularly useful for understanding group processes with the corporate board.

Additionally, some authors (Creswell et al., 2003a; Patton, 1990; Tashakkori & Teddlie, 1998) have enlarged the number of possible designs which can be used in studies combining a qualitative part and a quantitative one. They add another dimension (the stage of integration) to the already-mentioned implementation of data collection and priority. Thus, mixed methods typically refers to both data collection techniques and analysis, given that the type of data collected is so intertwined with the type of analysis that is used. However, we can unlink data collection and data analysis. Integration might occur within the research questions (e.g., both quantitative and qualitative questions are presented), within data collection (e.g., open-ended questions on a structured instrument), within data analysis (e.g., transforming qualitative themes into quantitative items or scales) or in interpretation (e.g., examining the quantitative and qualitative results for convergence of findings). Deciding on the stage or stages to integrate depends on the purpose to the research, the ease with which the integration can occur and the researcher’s understanding of the stages of research.

Tashakkori and Teddlie (1998), based on Patton (1990), establish a taxonomy of eight “mixed model studies” according to three dimensions: “confirmatory/exploratory investigation”, “quantitative/qualitative data/operations” and “statistical/qualitative data analysis and inference”. Thus, Tashakkori and Teddlie (1998) attempted to distinguish between mixed methods studies (mixed in the methods of study) and mixed model studies (mixed in more than the methods, including the questions and conclusions). However, following the recent developments in conceptualization of mixed methods, they abandon this distinction (Tashakkori & Creswell, 2007). Instead, they place mixed methods studies in two broad families of mixed studies and quasi-mixed studies. This quasi-mixed research identifies studies in which a serious integration of the findings/inferences does not occur (Teddlie & Tashakkori, 2006). Therefore, a key aspect in the definition of mixed methods research is integration (Bryman, 2007; Tashakkori & Creswell, 2007).

Teddlie and Tashakkori (2003) indicate what aspects a researcher can use to select the best-mixed methods design for his/her research project. It is obvious that the purpose of the study plays a large role. Moreover, typologies may be differentiated by the criteria that are used to distinguish among the research designs within them (e.g., priority and sequence). These criteria may be listed by the researcher, who then may apply the selected
criteria to potential designs, ultimately selecting the best research design for his/her study. In some cases, the researcher may have to develop a new mixed methods design because no one best design exists for research. Also, some designs may change over the course of the study. This might occur, for instance, if one type of data becomes more important as the study develops.

In our opinion, the practical relevance of any typology lies in the fact that it can help researchers to make decisions, which facilitate the design of their study. In any case, there are other perspectives available to assist researchers in the adoption of their main research decisions. Thus, Maxwell and Loomis (2003), from a more dynamic point of view, present a complementary approach to conceptualising mixed methods designs. Instead of presenting another typology of designs, they introduce an interactive model for research design in which its components (purpose, conceptual framework, research question, methods and validity) are components in a network or web rather than in a linear progression.

Along with the potential opportunity and contribution that these new designs and typologies can provide within the field of the RBV, it can equally prove interesting to identify purposes other than those mentioned in the preceding sections which have already been used in other fields and could consequently turn out to be useful in the RBV. Bryman and Bell (2003) present a wide variety of purposes in mixed methods research: (1) preparation: qualitative research facilitates quantitative research (providing hypotheses and aiding measurement – the in-depth knowledge of social contexts acquired through qualitative research can be used to inform the design of survey questions for structured interviewing and self-completion questionnaires); (2) preparation: quantitative research facilitates qualitative research (preparing the ground for qualitative research through the selection of people to be interviewed, or companies to be selected as case studies); (3) filling in the gaps (when the researcher cannot rely on either a quantitative or a qualitative method alone and must buttress his/her findings with a method drawn from the other research strategy, for example, qualitative research can help interpret and place in context the results of quantitative analysis); (4) static and processual features (whereas quantitative research can study the static features of a phenomenon, qualitative research can focus on more processual characteristics); (5) solving a puzzle (e.g., when quantitative findings are inconsistent with a hypothesis, a qualitative study could shed light).

Morgan (1998) refers to the purposes of complementary combinations of qualitative and quantitative research. Thus, he specifies the purposes that can be sought with each type of design. This author considers more practical
the four designs in which one of the methods is the dominant one and the implementation of data collection is sequential. First, in a qual→QUAN design, the previous qualitative study helps guide the data collection in a principally quantitative study and can generate hypotheses or develop content for questionnaires. Second, in a quan→QUAL design, the quantitative study helps guide the data collection in the qualitative study. Third, in a QUAN→qual design, the qualitative study helps evaluate and interpret results from a quantitative study. For example, the qualitative part can provide interpretation for poorly understood results, help explain outliers, etc. with in-depth interviews. Fourth, in a QUAL→quan design, the quantitative study helps evaluate and interpret results from a qualitative part, and it can generalize results to different samples or test elements of emergent theories.

An important challenge for mixed methods studies is the explicit clarification of several relevant aspects in the written report (Creswell et al., 2003a). Thus, in relation to data collection, the implementation decision calls for clearly identifying the core reasons for collecting both forms of data (quantitative and qualitative) in the first place, and understanding the important interrelationship between the quantitative and qualitative phases in data collection. Moreover, the choice of implementation strategy has consequences for the form of the final report. Thus, when two phases of data collection exist, the researcher can report the data collection process in two phases. The report may also include an analysis of each phase separately and the integration of information in the discussion or conclusion section of a study. In relation to priority, researchers need to make informed decisions about the weight or attention given to quantitative and qualitative research during all phases of their research. In the final written report, a project might be seen as providing more depth for one method than for the other through the length of discussions. Finally, the mixed methods researcher needs to design a study with a clear understanding of the stage or stages at which data will be integrated and the form this integration will take. The most typical case is the integration of the two forms of research at the data analysis and interpretation stages after quantitative data and qualitative data have been separately collected. For example, after collecting both forms of data, the analysis process might begin by transforming the qualitative data into numerical scores (e.g., themes or codes are counted for frequencies) so that they can be compared with quantitative scores (integrated quantitative data analysis). In another study, the analysis might proceed separately for both quantitative and qualitative data (quantitative data analysis and qualitative data analysis), and then the information might be compared and integrated in
the interpretation or discussion stage of the research. Less frequently is the integration at data collection. An example is the use of a few open-ended questions on a quantitative survey instrument. It is also possible for integration to occur earlier in the process such as in the question stage. In some studies, the researcher might set forth both quantitative and qualitative questions in which the intent is to both test some relationships among variables and explore some general questions. With the complex features often found in these designs, figures or visualizations can help the researcher to present the study.

Moreover, as said above, researcher must clearly identify the core reasons and rationale for collecting and combining both forms of data in a single study and make informed decisions about the weight or attention given to quantitative and qualitative research. Therefore, a mixed methods study must be competently designed and conducted, and as a consequence, researchers in the RBV must have the skills and training to carry out both quantitative and qualitative methods and take advantage of this type of research (Creswell, Tashakkori, Jensen, & Shapley, 2003b; Tashakkori & Teddlie, 2003c). Coupled with traditional statistics coursework, in-depth training concerning the discovery process (e.g., qualitative observational techniques or case studies) has the advantage of enabling researchers to think fully about what types of research questions are innovative, interesting and practically relevant (Currall et al., 1999).

An important challenge in mixed methods studies is the quality of this type of research (Bryman, 2006). The important point is that each method must be complete in itself, that is, all methods used must meet appropriate criteria for rigor (Morse, 1991). For instance, if qualitative interviews are conducted, then they must be conducted as if this method stands alone. Therefore, each part must satisfy its specific quality criteria. But additionally, a mixed methods study must satisfy some criteria linked to the mixed design. Caracelli and Riggin (1994) provide a list of mixed method quality criteria (e.g., a rationale for combining methods is provided or the use of mixed methods matches the stated purpose for combining the method types).

Some authors provide guidelines about how to carry out a mixed methods research. Creswell (1999) offers nine steps in conducting a mixed methods study: (1) determine if a mixed methods study is needed to study the problem; (2) consider whether a mixed methods study is feasible; (3) write both qualitative and quantitative research questions; (4) review and decide on the types of data collection; (5) assess the relative weight and implementation strategy for each method; (6) present a visual model;
(7) determine how the data will be analysed; (8) assess the criteria for evaluating the study and (9) develop a plan for the study.

Similarly, Teddlie and Tashakkori (2006) provide a seven step process for researchers selecting the best design for their projects: (1) the researcher must first determine if his/her research question require a monomethod or mixed methods design; (2) the researcher should be aware that a number of different typologies of mixed methods research designs and should know how to access details regarding them; (3) the researcher wants to select the best mixed methods design for his/her particular study and assumes that one of the published typologies includes the right design for the project; (4) typologies may be differentiated by the criteria that are used to distinguish among the research designs within them, and the researcher needs to know those criteria; (5) these criteria should be listed by the researcher, who may then select the criteria that are most important for the particular study he/she is designing; (6) the researcher then applies the selected criteria to potential designs, ultimately selecting the best research design and (7) in some cases, the researcher may have to develop a new mixed methods design, because no one best design exists for his/her research project.

Regarding the last step, Teddlie and Tashakkori (2006) point out that mixed methods designs have an opportunistic nature. Thus, in many cases, a mixed methods research study may have a predetermined research design, but new components of the design may evolve as researchers follow up on leads that develop as data are collected and analysed. These opportunistic designs may be different from those contained in previously published typologies. The point is for the researcher to be creative and not be limited by the existing designs. A design may emerge during a study in new ways, depending on the conditions and information that is obtained. Thus, a tenet of mixed methods research is that researchers should mindfully create designs that effectively answer their research questions (Johnson & Onwuegbuzie, 2004).

Hanson, Creswell, Plano Clark, Petska, and Creswell (2005) offer some recommendations for designing, implementing and reporting a mixed methods study. Thus, they recommend that researchers attend closely to design and implementation issues, particularly to how and when data are collected (e.g., concurrently or sequentially). The study’s purpose plays an important role here. They also recommend that researchers familiarize themselves with the analysis and integration strategies used in the published mixed methods studies. Moreover, because mixed methods studies require a working knowledge and understanding of both quantitative and qualitative methods, and because they involve multiple stages of data collection and analysis that frequently extend over long periods of time, they recommend
that researchers work in teams. Moreover, in preparing a mixed methods manuscript, they recommend that researchers use the phrase “mixed methods” in the titles of their studies, and that, early on, researchers foreshadow the logic and progression of their studies by stating the study’s purpose and research questions in the introduction. Clear, well written purpose statements and research questions that specify the quantitative and qualitative aspects of the study help focus the manuscript. Additionally, these authors recommend that, in the introduction, researchers explicitly state a rationale for mixing quantitative and qualitative methods and data (e.g., to triangulate results, to develop or improve one method with the other, to extend the study’s results). Another recommendation is that, in the methods, researchers specify the type of mixed methods research design used. By doing this, the field will be able to build a common vocabulary and shared understanding of the different types of designs available.

**Contributions to Mixed Methods Research**

I would like to point out two main contributions of this literature review of RBV mixed methods studies to the general field of mixed methods research. First, qual→QUAN design dominates, and this fact can be due to the youth of this perspective. Several aspects of the quantitative part may be improved through the use of a previous qualitative phase. As said above, this qualitative part may help to develop or extend theory, determine the adequate level of analysis, identify the industry-specific resources and competences as well as the dependent variables, improve the measurement instrument of the quantitative phase and enhance the interpretation of quantitative results. An important issue here is the research question. As said previously, Creswell (1999) pointed out that in conducting a mixed methods research it is necessary that you write both qualitative and quantitative research questions. However, in my opinion, mixed methods research also provides opportunities to better answer a single research question, even if that question are linked to a pure approach. Thus, qual→QUAN studies in the RBV, with a quantitative research question and the qualitative part providing helpful information indicated above, are examples about this issue.

Second, some authors have pointed out that sequential mixed methods designs are usually illustrated as having only two parts and phases, but they could be more complex, involving three or more parts or strands (Johnson & Onwuegbuzie, 2004; Teddlie & Tashakkori, 2006). However, few examples are provided. Previously, I pointed out the study by Currall et al. (1999).
Additionally, our literature review in the field of the RBV provides other examples. Specifically, although in the review the study by Hoetker (2005) was classified as a QUAN+qual design (the author uses quantitative and qualitative data to capture a more complete, holistic and contextual portrayal of its phenomenon of study), the author points out that the understanding of the industry draws upon discussions with numerous industry informants. Then, this study can be considered as a qual→(QUAN+qual) design. Dyer and Hatch (2006) is another example. This study was classified in the literature review as a QUAN→QUAL design (survey about a quantitative objective and then interviews to explore a qualitative objective). However, it can also be considered as a qual→QUAN→QUAL design, because the authors pointed out that previous interviews with plant managers and their top assistants at 10 suppliers help them to develop the survey of the quantitative phase.

Thomke and Kuemmerle (2002), in their study about asset accumulation in pharmaceutical drug discovery, pointed out in the research design that data on the accumulation of chemical libraries and the related asset stock were collected in three stages. In the first stage, they followed a grounded research approach, which involved interviews with experts involved in drug developments. This led to the identification of chemical libraries and related molecule screening and refining assets as a bundle of interdependent assets. The field research also resulted in a case study. Interviews and other qualitative data collected at this first stage allowed them to build a deeper understanding of the subtle technology and process issues. They use this in-depth understanding to design a questionnaire (second stage) that was used to collect data on the build process of chemical libraries, its impact on discovery output and the role of new combinatorial chemistry and high-throughput screening. I classified this study as a QUAL→QUAN, based on my opinion that qualitative and quantitative parts have similar importance and because both data were used in findings section of the study. However, in the research design, the authors added two additional aspects: (1) that after all firms had returned the questionnaires, they followed up with interviews to make sure that the questions had been understood as intended and to identify potential measurement problems; (2) in a third stage, they focused on the trading and adoption of externally developed technology assets, and as proxies for such activities they compiled a list of and analysed all publicly announced alliances from 1986 to 1996 involving combinatorial chemistry and high-throughput technologies. In this third stage, they used several data sources and collected mainly qualitative data for each alliance (intended purpose, timing, partners involved and objectives). This information was
also used in the findings section. Therefore, this study could be considered as a QUAL → QUAN → qual → QUAL design.

**CONCLUSIONS**

This chapter has tried to contribute to the progress of empirical research on the RBV providing several ideas about the application of mixed methods designs on this field. First, several issues about mixed methods to make clear what the literature already tells us about this type of research have been examined. Second, I have carried out a review about the use of mixed methods in the RBV, pointing out the main characteristics of an illustrative example in each type of design. Finally, some implications for future research have been examined. Specifically, it has been showed a comparison between RBV monomethod and mixed methods studies, some suggestions and recommendations about the use of mixed methods based on previous experience in other fields, and contributions of RBV mixed methods in the general field of mixed methods research. **Table 2** shows a summary of guidelines for mixed methods research that have been pointed out in previous sections.
I would like to indicate that I agree with the “paradigm of choices” emphasized by Patton (1990). A paradigm of choices rejects methodological orthodoxy in favour of methodological appropriateness as the primary criterion for judging methodological quality. Thus, this paradigm of choices recognizes that different methods are appropriate for different situations. The predominance of more quantitative-based methodological tools in the development of the strategic management and the RBV does not mean that these tools are applicable to all research questions. The research question and context should dictate the choice of the appropriate research methods. However, we must also take into account that the knowledge about mixed methods research can stimulate a researcher to better define and analyse innovative problems and research questions in strategy and RBV research. Mixing methods therefore offers enormous potential for exploring new dimensions (Mason, 2006).

To our knowledge, this is the first study to carry out a review of the presence of mixed methods in a specific strategic area, specifically the RBV. Hopefully, this review of RBV empirical studies which have used mixed methods designs along with the ideas offered for the application of mixed methods studies may favour progress both in the future research on the RBV and in the studies devoted to mixed methods research.

ACKNOWLEDGEMENTS

A first version of this manuscript was presented at the First International Conference on Research Methods, co-sponsored by Research Methods Division of the Academy of Management and ISEOR, held in Lyon (France) in 2004. I am grateful to the participants who discussed the paper in this conference. I am also grateful to Donald Bergh and David Ketchen for helpful comments and suggestions during the development of this chapter.

REFERENCES


This page intentionally left blank
TOWARD GREATER INTEGRATION OF THE RESOURCE-BASED VIEW AND STRATEGIC GROUPS RESEARCH: AN ILLUSTRATION USING RANDOM COEFFICIENTS MODELING

Jeremy C. Short

ABSTRACT

The resource-based view (RBV) of the firm focuses on how firm-level assets and capabilities influence firm performance. Scholars have noted the need for studies grounded in the RBV to account for the role of the strategic group level, but uncertainty remains about how to do so. Random coefficients modeling (RCM) provide an appropriate technique to integrate these two levels of analysis, but its use has been limited in strategic management research to date. I review research integrating firm and strategic group levels and provide a roadmap for future research seeking to integrate these two levels’ influences on firm performance, and use RCM to illustrate the effects of firm resources on performance under three depictions of the strategic group level culled from strategic management research. Findings suggest that interpretations about the efficacy of resources’ influence on performance...
vary considerably across methodological specification. Next, I use RCM to illustrate how strategic management researchers can further integrate the firm and group levels by demonstrating how variables at the group level of analysis may interact with firm-level characteristics. I conclude with suggestions for future research using RCM to integrate the strategic group into multilevel studies predicting firm performance.

Understanding the determinants of firm performance is a central goal of the strategic management field. The resource-based view (RBV) of the firm is perhaps the most prominent perspective aimed at building knowledge about performance. The RBV examines factors at the firm level of analysis that have the potential to influence outcomes. The central tenet of the RBV is that superior firm performance is a function of the firm’s ability to both accumulate and deploy scarce assets and capabilities (Barney, 1991). Firm resources will lead to sustained competitive advantage when they are valuable, rare, without substitutes, and bundled in a manner so that the firm’s resources, and thus strategies, are inimitable by current and future competitors (Barney, 1991).

A challenge for RBV studies is how to incorporate strategic groups. Strategic groups research focuses on sets of firms that follow similar strategies, and this research stream is often concerned with understanding how group membership may influence performance. This literature is based on the premise that profits differ systematically among sets of firms within an industry because of market factors and similar asset profiles that are common to groups of firms (Porter, 1979). Indeed, a number of studies have found evidence for this premise (e.g., Dess & Davis, 1984; Reger & Huff, 1993). In addition, a meta-analysis found that the group level plays a significant role in performance (Ketchen, et al., 1997), and empirical research has demonstrated that the group level is significant in determining performance when examined in conjunction with firm and industry effects (Fox, Srinivasan, & Vaaler, 1997; Short, Ketchen, Palmer, & Hult, 2007). Thus, the strategic group is an important level of analysis.

Given the two areas’ prominence, there have been calls to integrate them in order to create a more complete understanding of the antecedents to performance (Mahoney & Pandian, 1992; Rouse & Daellenbach, 1999). Before integration begins in earnest, it is important for researchers to understand how their decisions about how to incorporate data from both levels of analysis may influence findings and apply appropriate analytical techniques to test the effects of phenomena at multiple levels of analysis.
Random coefficients models (RCM), also known as hierarchical linear models, multilevel linear models, mixed-level models, and variance components models, provide a useful tool to aid in the integration of firm and strategic group levels of analysis. Such techniques provide a number of advantages over traditional estimation techniques such as regression (Hofmann, 1997), and the application of RCM was greatly aided by developments in the mid-1980s that made the estimation of such models more feasible (Hofmann, Griffin, & Gavin, 2000). Recently, researchers have begun to incorporate RCM techniques to examine the determinants of firm performance (e.g., Misangyi, Elms, Greckamer, & Lepine, 2006; Short, Ketchen, Bennett, & Du Toit, 2006). Unfortunately, the application of such techniques to test firm and strategic group level phenomena has been minimal (Short et al., 2007); thus, RCM continues to be an underutilized tool to incorporate the strategic group level of analysis in strategic management research.

To illustrate a useful empirical technique to integrate the strategic group level of analysis, this paper offers three contributions to the literature. The first is providing a synthesis of previous research addressing the firm and strategic group levels. The second contribution is outlining a roadmap for future research attempting to integrate firm and strategic group influences on performance. The third contribution is demonstrating how RCM can be used to test the influences of the firm, strategic group, and industry levels of analysis. To illustrate the uses of RCM, I use the Hierarchical Linear Modeling 6.0 software package to assess the effects of different methodological assumptions that arise from researchers’ choices when applying RCM; I then use HLM to provide two illustrations involving tests of variables at the firm and group levels of analysis. I conclude with suggestions for future research seeking to apply a multilevel approach to integrate the firm, strategic group, and industry levels of analysis.

**INTEGRATING THE FIRM AND STRATEGIC GROUP LEVELS**

A variety of works in strategic management have noted the importance of multiple levels of analysis when seeking to understand the determinants of firm performance. The early work of Porter (1979, 1980), for example, questioned conventional economics logic that argued that superior firm performance was primarily a function of phenomena at the industry level of analysis. Porter contended that firm level differences, as well as structural
differences within industries caused by strategic groups, influenced firm performance outcomes. As the strategic management field has developed over the last three decades, research has embraced conceptual and methodological advancements that incorporate multiple levels of analysis including the firm, strategic group, and industry levels (Hoskisson, Hitt, Wan, & Yiu, 1999).

The emerging multilevel approaches to strategic management overlap with a trend in organizational research in general that emphasizes a meso paradigm which recognizes the importance of linkages among multiple levels of analysis (House, Rousseau, & Thomas-Hunt, 1995). For example, a number of studies in strategic management have tested the influences of the firm and industry levels of analysis on firm performance (e.g., Rumelt, 1991). Such an approach traces its roots to systems theory, a concept which suggests that understanding organizational phenomena is best approached by viewing the actions of individual actors such as a single firm within the context of a broader system such as an industry or economy (Ashmos & Huber, 1987; Katz & Kahn, 1978).

Despite the multilevel focus in strategy research in general, integration of one level of analysis, the strategic group, has been slow in strategic management research (Short, Palmer, & Ketchen, 2003). Strategic groups research originated in Hunt’s (1972) dissertation, which found that firms in the US appliance industry could be meaningfully grouped based on differences in vertical integration, product diversification, and product differentiation. Firms within the industry faced different competitive threats depending on group membership. Arising from this early work, strategic groups research is based on the premise that profits differ systematically among groups of firms within an industry because of market factors and similar asset profiles that are common to groups of firms (Porter, 1979).

Difficulties associated with capturing strategic group effects may be one reason strategic groups research has been hindered historically and difficult to integrate into research involving firm and industry levels of analysis. Construct measurement has been a challenge in the relatively young strategic management field in general (Boyd, Gove, & Hitt, 2005b), and the strategic group level has been particularly difficult to isolate. Although strategic groups are theorized to be naturally occurring subsets of firms within an industry, researchers have often measured strategic groups through cluster analysis or other methods of placing firms into groups. As a consequence, group characteristics have often been measured through aggregates of firm level measures. Consequently, assessments of the influence of strategic groups on performance have often relied on ANOVA
techniques that assessed differences in group level performance. This analytical strategy led to concerns that group level differences may not adequately or accurately inform researchers, scholars, and practitioners about the influence of group level phenomena on firm level performance (Barney & Hoskisson, 1990).

Despite the inherent challenges, a number of previous studies have successfully incorporated both firm and strategic group levels (see Table 1). Lawless et al. (1989) answered some of the criticisms of previous scholars and found performance differences both within and between strategic groups. McNamara et al. (2003) used hierarchical linear modeling (HLM) to

**Table 1.** Research Examining Firm and Strategic Group Level Influences on Performance.

<table>
<thead>
<tr>
<th>Primary Research Question</th>
<th>Sample Characteristics</th>
<th>Statistical Technique</th>
<th>Example of Published Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there firm level performance differences within strategic groups?</td>
<td>Multiple industries, strategic groups, and firms</td>
<td>OLS regression with dummy variables to assess group differences</td>
<td>Lawless, Bergh, and Wilsted (1989)</td>
</tr>
<tr>
<td>How much variance exists within and between strategic groups?</td>
<td>Single industry</td>
<td>Hierarchical linear modeling</td>
<td>McNamara, Deephouse, and Luce (2003)</td>
</tr>
<tr>
<td>How much variance in performance do the firm, strategic group, and industry levels explain?</td>
<td>Multiple industries, strategic groups, and firms</td>
<td>Hierarchical linear modeling</td>
<td>Short et al. (2007)</td>
</tr>
<tr>
<td>How much does strategic group membership influence performance, controlling for the effects of firm resources?</td>
<td>Single industries with evidence of strategic groups</td>
<td>OLS regression with dummy variables to assess strategic group membership</td>
<td>Nair and Kotha (2001)</td>
</tr>
<tr>
<td>How does strategic group membership shape the long-run competitive dynamics among firms</td>
<td>Single industries with evidence of strategic groups</td>
<td>Cointegration analysis</td>
<td>Nair and Filer (2003)</td>
</tr>
<tr>
<td>Does strategic group membership moderate the relationship between firm resources and performance?</td>
<td>Single industries with evidence of strategic groups</td>
<td>OLS regression with dummy variables to assess strategic group membership and interaction terms to test moderation</td>
<td>Short et al. (2002)</td>
</tr>
</tbody>
</table>
partition the variance into within and between group components in a single analysis. Researchers have also noted that the strategic group level of analysis explains unique variance in performance not associated with firm or industry levels (Fox et al., 1997; Short et al., 2007). Other empirical efforts to integrate the firm and strategic group levels of analysis have tested the importance of group membership above and beyond the influences of firm resources (Nair & Kotha, 2001). More recent studies have examined the influences of group membership on competitive dynamics among firms (Nair & Filer, 2003) and the moderating influences of group membership on the relationship between firm level resources and performance (Short, Palmer, & Ketchen, 2002).

Further integration of multilevel theory and techniques highlight a number of limitations of previous research efforts and suggest opportunities for future research integrating the firm and strategic group levels. Conceptually, multilevel researchers have noted the importance of general systems theory as a holistic view that would provide a multilevel perspective in organizational science. The hope is to identify principles that would enhance our understanding of phenomena by embracing multiple levels of analysis (Kozlowski & Klein, 2000). While systems theory has been an important influence in strategic management research historically (Hoskisson et al., 1999), this framework has received little attention as a conceptual lens for studies integrating the firm and strategic group levels of analysis. Perhaps this lack of integration is partially due to the fact that strategy is a relatively young field (Boyd, Finkelstein, & Gove, 2005a). As a consequence, more scholarship may be devoted to promoting and defending particular academic camps rather than integrating conceptual frameworks. Thus, it may come as little surprise that previous efforts to incorporate firm and strategic group levels have often focused on the isolation of firm and group level effects (e.g., Fox et al., 1997; Nair & Kotha, 2001), rather than any attempt to truly integrate these two levels of analysis. Empirically, researchers incorporating firm and group levels have often integrated group membership via dummy variables that represent groups within the context of a single industry (Nair & Kotha, 2001; Short et al., 2002). Although this specification provides a critical first step, it makes tests of measures of group characteristics (i.e., variables at the group level of analysis) difficult to integrate since the number of strategic groups in a single industry (generally less than ten) is unlikely to have sufficient sample size to detect small or medium effects. Studies testing a single industry context also limit our understanding about the generalizability of group characteristics, inhibiting insights as to the importance of strategic groups across multiple industries.
Future research incorporating the firm and strategic group levels of analysis would benefit from recent advancements in empirical techniques that are appropriate to test variables at multiple levels of analysis. Particularly, RCM techniques such as HLM provide the ability to test main and moderating effects at the group level on relationships at the individual level (Hofmann, 1997). Such techniques allow for a more thorough integration of nested relationships such as firms within strategic groups within industries. Although researchers have used HLM to test firm and group effects (McNamara et al., 2003), and firm and industry effects (Misangyi et al., 2006; Short et al., 2006) the advantages of such techniques to fully integrate the strategic group level of analysis remain underutilized in strategic management research. Table 2 displays a number of research questions that could use RCM/HLM to provide a contribution to knowledge. These suggestions differ from previous efforts in a number of ways. Most notably, many of these suggestions test the moderating effects of group level characteristics, rather than simply group membership alone. Because the role of a firm’s competitive environment is critical when understanding firm performance (Morrow, Johnson, & Busenitz, 2004), a number of these research questions involve the incorporation of higher levels of analysis such as industry level effects. In the following sections, I provide two empirical examples to illustrate why and how future research on the firm and strategic groups levels could begin to capitalize on RCM/HLM to test multilevel models.

USING RANDOM COEFFICIENTS MODELING TO INTEGRATE THE FIRM AND STRATEGIC GROUP LEVELS

A review of research integrating the firm and strategic group levels suggests a number of fruitful research questions that would provide a more thorough integration of these two levels than previous efforts. In this section, I introduce the merits of RCM as an appropriate analytical technique to test a number of these research questions. In brief, RCM techniques are attractive to strategic management researchers because they allow for analysis of variance at multiple levels of analysis (traditionally accomplished via ANOVA or variance decomposition software), as well as the introduction of independent variables at each level of analysis (traditionally tested using regression) using a single analytical technique. Further, such techniques allow for a multilevel modeling of moderating effects at higher levels of analysis (Hofmann, 1997). My goal is to illustrate the use of RCM, as well
as point out the empirical challenges researchers will likely face when testing firm and strategic group levels of analysis in multilevel strategy research.

Sample

I drew my sample from the COMPUSTAT database. I restricted the sample to firms that operate in a single business segment to eliminate statistical noise that would occur if I attempted to measure diversified firms operating
in multiple industries (cf. Mauri & Michaels, 1998). A lagged structure was used to improve the ability to make causal inferences (cf. Palmer & Wiseman, 1999). Firm resources and strategic group traits were measured with data taken from 1991 to 1995, and firm performance variables were averaged from 1993 to 1997. The decision to use five years of data for each measure was driven by the need to provide a stable basis for comparing competitive strategies (cf. Miles, Snow, & Sharfman, 1993) and to provide a stable measure of firm outcomes (cf. Keats & Hitt, 1988). Three years of data overlap were chosen because some resources may have immediate effects on performance, while others may take a number of years before their performance affects are fully realized (cf. Palmer & Wiseman, 1999).

The sample was drawn from several manufacturing industries to enhance generalizability. In addition to the single business requirement, the sample was further restricted to industries with a minimum of 45 firms to have the statistical power needed to detect medium effects (Ferguson & Ketchen, 1999). Specifically, I sampled from 12 four-digit SIC industries where competition is dominated by single business firms. This selection process resulted in a total sample of 1,163 firms.

**Firm Resource and Performance Measures**

To examine firm resources, I relied on the classification of Chatterjee and Wernerfelt (1991), which places firm resources into three categories: physical, intangible, and financial. To provide a measure of physical resources that was meaningful across the 12 industries included in the sample, I used capital intensity, defined as capital expenditures divided by sales. A firm that makes a consistent commitment to capital expenditures is continually building their property, plant, and equipment. To operationalize intangible resources, I used the number of patents granted to the firm between 1991 and 1995, available from the CASSIS database from the Patent and Trademark Office of the U.S. Department of Commerce (Penner-Hahn, 1998). Because patent protection provides the owner with exclusive rights to make, use, and sell the patented invention for more than a decade, a patent is an intangible resource that can be used by firms to achieve a sustained competitive advantage (Hall, 1992; Markman, Espina, & Phan, 2004). Available financial resources provide the means for achieving strategic flexibility (i.e., slack) that can enhance organizational performance (Greenly & Oktemgil, 1998). Following Chatterjee and Wernerfelt (1991), I used the current ratio to measure financial resources.
To indicate performance, I used return on assets (ROA). This measure has been used as an indicator of performance in several multilevel studies of firm performance (e.g., Keats & Hitt, 1988; Mauri & Michaels, 1998; Rumelt, 1991; Short et al., 2006), and has often been the sole measure of firm performance in many of these studies (e.g., McGahan & Porter, 1997; Rumelt, 1991). Correlations among firm level variables and performance are displayed in Table 3.

**Table 3.** Correlations among Firm-Level Variables.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Return on assets</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Capital intensity</td>
<td>-.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Patents</td>
<td>.07*</td>
<td>-.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Current ratio</td>
<td>-.15**</td>
<td>.10**</td>
<td>.13**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. R&amp;D intensity</td>
<td>-.39**</td>
<td>.09**</td>
<td>-.01</td>
<td>.31**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6. Trademarks</td>
<td>.18**</td>
<td>-.06*</td>
<td>.17**</td>
<td>-.02</td>
<td>-.09**</td>
<td>1</td>
</tr>
</tbody>
</table>

*p < .05.

**p < .01.

Measuring Strategic Groups

Defining strategic groups has been a challenge both conceptually as well as empirically. Porter (1980, p. 129) defines a strategic group as a “group of firms in an industry following the same or a similar strategy”. Cool and Schendel (1987, p. 1106) apply a more general definition and describe strategic groups as “firms competing within an industry on the basis of similar resource combinations of scope and resource commitments.” Despite differences in semantics, researchers agree that strategic groups are naturally occurring subsets of firms that are more homogeneous in actions than other industry incumbents (Cool & Schendel, 1988).

Although strategic group researchers view these groups of firms as naturally occurring groups, assignment of group membership has posed a number of measurement difficulties.

Much of the empirical research using strategic groups has relied on cluster analysis, an analytical technique that forms groups based upon similarities of firm level variables (Ketchen & Shook, 1996). The logic of cluster analysis matches the conceptualizations of strategic groups as groups of firms with similar asset profiles. Although this approach has practical as well as
conceptual appeal, the aggregation of firm level variables to form groups opens this technique to criticism from multilevel scholars who argue that researchers should carefully measure data at the appropriate level of analysis (Klein, Dansereau, & Hall, 1994; Rousseau, 1985).

In addition to difficulties associated with aggregating firm level variables to measure group phenomena, early work using cluster analysis was also criticized for a lack of theoretical basis when selecting the dimensions along which to cluster; thus, the groups of firms derived from cluster analysis is often sample (i.e., industry) specific (Ketchen & Shook, 1996). To ameliorate this concern, however, some researchers have relied on deductively defined groups where the variables used to define groups are theoretically viable at multiple levels of analysis and the results have an interpretation across industries (Ketchen, Thomas, & Snow, 1993; Bantel, 1998).

To illustrate an examination of multiple strategic groups within a number of industries, I rely on such a deductive approach for its superior generalizability (Bantel, 1998). The deductive approach used here relies on two theoretical perspectives at the heart of organizational analysis – strategic choice (Miles & Snow, 1978) and organizational ecology (Hannan & Freeman, 1977). The basic premise is that a firm’s strategy varies on two independent competitive dimensions central to both theories (Zammuto, 1988). The first relates to a firm’s competitive advantage, expressed as the ability to exploit new opportunities. Being first to market, or otherwise exploiting a new market quickly, may accomplish this. The second dimension focuses on breadth of operations. This dimension includes number of distributors, geographic or product scope, and market growth/share goals (Bantel, 1998). A strong competence in only one of these dimensions is needed for sustained competitive advantage.

Combining these two dimensions results in four distinct quadrants that closely parallel the Miles and Snow (1978) typology. The first quadrant is represented by defenders/K-specialists, who focus on existing opportunities in a narrow domain. The second quadrant encompasses entrepreneurs/r-specialists, who pursue existing opportunities in a narrow domain. Analyzers/K-generalists, efficiently exploit existing opportunities in a broad domain. Finally, prospectors/r-generalists, pursue new opportunities in a broad domain. Fig. 1 illustrates the deductive groups used to cluster firms for each of the 12 industries included in the sample.

I used two measures to cluster strategic groups along each of the competitive dimensions. To measure competitive advantage, I used R&D intensity (cf. Bantel, 1998). A firm that makes a significant, consistent investment to R&D has the capability to create an innovation or be an early
follower (Schoenecker & Cooper, 1998). I measured R&D intensity as the average R&D expenditure divided by sales for the years 1991–1995 (Bierly & Chakrabarti, 1996). To measure breadth of operations, I used the number of trademarks the firm holds. Trademarks proxy for operations breadth because firms with a large number of trademarks are likely to be involved in the production of numerous products, services, or devices (Hall, 1992). Conversely, firms that hold few trademarks are more likely to focus their operations on a narrow niche market. Thus, trademarks capture a number of unobservables associated with competitive scope.

A two-stage clustering procedure was used to cluster the firms in the analysis. A two-stage process is valuable because it increases the validity of cluster solutions (Ketchen & Shook, 1996). This procedure first uses hierarchical clustering to determine the number of groups and their cluster centroids (i.e., Ward’s method) and then uses the results as the starting point for a nonhierarchical clustering (i.e., K-means).

The use of a deductively defined framework for clustering strategic groups resulted in four groups for each of the industries in the sample. Thus, the final sample included 1,163 firms in 48 strategic groups in 12 industries. Correlations among group level variables are displayed in Table 4.

---

**Fig. 1.** Strategic Types Combining the Organizational Ecology and Strategic Choice Perspectives. Adapted from Zammuto (1988).
I use the three level HLM software package to illustrate the effects of firm resources on performance under three depictions of the strategic group level culled from strategic management research. The HLM 6.0 software provides tests following the assumptions of OLS regression, as well as under different assumptions germane to random coefficients models (Raudenbush, Bryk, Cheong, & Congdon, 2000). The three level model tests the influences of firm performance based on the firm level of analysis (level-1), strategic group level of analysis (level-2), and industry level of analysis (level-3), and in the following sections I highlight differences in empirical specification. Each of these specifications has conceptual grounding in the strategic management literature.

The first specification suggests that the context of higher levels of analysis (such as the strategic group and industry) has little effect on how firm resources influence performance. This specification allows me to demonstrate the advantages of RCM techniques for multilevel research in strategic management, as I briefly compare the assumptions of RCM techniques with those of ordinary least squares (OLS) regression. Under the assumptions of OLS regression, testing the influence of a firm level resource (e.g., capital intensity) on a single firm performance variable (e.g., ROA) would assume that the value of a specific resource is invariant across strategic groups, and even across industries. For example, testing the relationship between an individual level independent variable on an individual level dependent variable would take the following familiar form:

\[ Y = b_0 + b_1 X_1 + e \]

Thus, a specification using OLS regression assumes that influences at higher levels of analysis (such as strategic groups and/or industries) have negligible effects on performance. Conceptually, this specification is problematic

**Table 4.** Correlations among Group-Level Variables.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. R&amp;D intensity</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Trademarks</td>
<td>-.23</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3. Group size</td>
<td>-.20</td>
<td>-.30*</td>
<td>1</td>
</tr>
</tbody>
</table>

*p < .05.

**Illustration Testing the Effects of Firm Resources under Differing Assumptions Concerning Group Level Influences Using RCM**

I use the three level HLM software package to illustrate the effects of firm resources on performance under three depictions of the strategic group level culled from strategic management research. The HLM 6.0 software provides tests following the assumptions of OLS regression, as well as under different assumptions germane to random coefficients models (Raudenbush, Bryk, Cheong, & Congdon, 2000). The three level model tests the influences of firm performance based on the firm level of analysis (level-1), strategic group level of analysis (level-2), and industry level of analysis (level-3), and in the following sections I highlight differences in empirical specification. Each of these specifications has conceptual grounding in the strategic management literature.

The first specification suggests that the context of higher levels of analysis (such as the strategic group and industry) has little effect on how firm resources influence performance. This specification allows me to demonstrate the advantages of RCM techniques for multilevel research in strategic management, as I briefly compare the assumptions of RCM techniques with those of ordinary least squares (OLS) regression. Under the assumptions of OLS regression, testing the influence of a firm level resource (e.g., capital intensity) on a single firm performance variable (e.g., ROA) would assume that the value of a specific resource is invariant across strategic groups, and even across industries. For example, testing the relationship between an individual level independent variable on an individual level dependent variable would take the following familiar form:

\[ Y = b_0 + b_1 X_1 + e \]

Thus, a specification using OLS regression assumes that influences at higher levels of analysis (such as strategic groups and/or industries) have negligible effects on performance. Conceptually, this specification is problematic
because if groups and/or industries are known to influence performance, this specification violates the independence of observations assumption that underlies traditional statistical techniques (Hofmann, 1997; Raudenbush & Bryk, 2002).

In contrast to OLS regression, RCM techniques address a number of potential measurement difficulties associated with tests of firms within higher levels of analysis such as strategic groups and industries. RCM techniques proceed with a nested structure of firms. For example, I test firms nested within strategic groups nested within industries. This specification produces a different variance for each level for the factor measured at that level. Thus, if the research goal is to measure phenomena at different levels of analysis, an OLS regression equation that predicts performance at only one level is inappropriate because strategic groups will be composed of multiple firms and the group characteristics will not vary for the firms in each group (thus, they lack independence and the associated error terms also lack independence). Additionally, estimating firm and group effects in the same equation assumes that the firm and group characteristics are from a simple random sample, which cannot be true because group characteristics are by definition invariant for all the firms in the same group. Another problem is that firms in the same strategic group or industry tend to be more alike than do firms in other strategic groups or industries; thus, the firms themselves are not from a random sample. The variance between firms may not be constant, but instead vary between groups and industries. Therefore, OLS regression misestimates characteristics in tests that involve multiple levels of equations. RCM techniques are able to ameliorate such concerns and correct for these errors, if present.

Using RCM, the firm, strategic group, and industry levels of analysis are all represented. The level-1 (firm level of analysis) model represents the performance for each firm as a function of a strategic group mean plus random error using the following equation:

\[
\text{Performance}_{ijk} = \pi_{0jk} + e_{ijk}
\]

where \(\text{Performance}_{ijk}\) is the average performance for a single dependent variable (i.e., ROA) of firm \(i\) in strategic group \(j\) and industry \(k\); \(\pi_{0jk}\) is the mean performance of strategic group \(j\) in industry \(k\); \(e_{ijk}\) is a random “firm effect” that measures the deviation of firm \(ijk\)’s score from the strategic group mean. These effects are assumed normally distributed with a mean of 0 and variance \(\sigma^2\). The subscripts \(i, j,\) and \(k\) denote firms, strategic groups, and industries where there are \(i = 1,2,\ldots, n_{jk}\) firms within strategic group \(j\) in
industry $k; j = 1,2,\ldots,J_k$ strategic groups within industry $k$; and $k = 1,2,\ldots,K$ industries.

The level-2 model examines each strategic group mean, $\pi_{0jk}$ as an outcome varying randomly around some industry mean using the following formula:

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

where $\beta_{00k}$ is the mean performance in industry $k; r_{0jk}$ is a random “strategic group effect,” that is, the deviation of strategic group $jk$’s mean from the industry mean. These effects are assumed normally distributed with a mean of 0 and variance $\tau_\pi$. Within each of the $K$ industries, the variability among strategic groups is assumed the same.

The level-3 model represents the variability among industries. The industry mean, $\beta_{00k}$, varies randomly around a grand mean as presented in the following formula:

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

where $\gamma_{000}$ is the grand mean; $u_{00k}$ is the random “industry effect,” that is, the deviation of industry $k$’s mean from the grand mean. These effects are assumed normally distributed with a mean of 0 and variance $\tau_\beta$.

Thus, to test the second specification of the effects of a single independent variable on a single dependent variable, the following set of equations are estimated.

$$\text{Performance}_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{capital intensity})_{ijk} + e_{ijk}$$

$$\pi_{0jk} = \beta_{00j} + r_{0ij}$$

$$\pi_{1jk} = \beta_{10j} + r_{1ij}$$

$$\beta_{00k} = \gamma_{000} + U_{00k}$$

When using RCM techniques, researchers should be aware that the method of scaling (i.e., centering) level-1 predictor variables can influence results significantly and should be based on theoretical paradigm and research question (Hofmann & Gavin, 1998). In the previous example, I use grand mean centering, where the grand mean of the level 1 variable is subtracted from each individual’s score. This specification yields equivalent models to raw-metric approaches, where level 1 predictors are used in their original form (Hofmann et al., 2000).
The third and final specification, the application of group mean centering, provides an empirical scaling option that may be particularly attractive for researchers testing the influences of firm resources within strategic groups. In the sample of multiple firms nested within multiple strategic groups, group mean centering specifies that the effect of a specific firm resource (e.g., capital intensity) is centered (i.e., subtracted) from each firm and thus the subtracted mean is not a constant, and varies based on strategic group membership. The assumption of this specification has considerable influence when integrating the strategic group level of analysis into tests of firm resources. Specifically, group mean centering would suggest that the efficacy of firm resources varies based on the group where a firm is a member. Thus, the value of a specific resource is based upon accumulation of the resource compared to other members of the same strategic group. This specification is supported by strategic groups research that notes the importance of strategic groups as referents (Fiegenbaum & Thomas, 1995), and notes strategic differences between ‘core’ strategic group members that closely follow a strategic group recipe and ‘secondary’ members that implement the strategic group recipe less consistently than core firms (Reger & Huff, 1993; McNamara et al., 2003). Testing of this specification would take the following form:

\[
\text{Performance}_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{capital intensity}_{\text{group centered}})_{ijk} + \epsilon_{ijk}
\]

\[
\pi_{0jk} = \beta_{00j} + r_{0ij}
\]

\[
\pi_{1jk} = \beta_{10j} + r_{1ij}
\]

\[
\beta_{00k} = \gamma_{000} + U_{00k}
\]

Table 5 displays the firm resource variables under the assumptions of OLS regression, as well as grand mean and group mean centering options using HLM. Although RCM/HLM approaches can effectively test multiple variables at each level of analysis, I test each firm level variable separately to illustrate the effects of centering options and to avoid multicollinearity. As shown in Table 5, the empirical and theoretical assumptions have a dramatic influence on the tests of firm resource effects on ROA. Using OLS assumptions, capital intensity was not related to ROA, number of patents had a positive relationship with ROA, and current ratio had a negative effect. Using HLM “grand mean” centering, capital intensity had a significant negative relationship with ROA, while number of patents and
the current ratio were not significant. Using HLM “group mean” centering, capital intensity and number of patents were not significantly related to ROA, but current ratio had a significant positive relationship to ROA.

These findings suggest fundamental substantive differences will be detected based on use as well as specification of RCM. For example, under the assumptions of OLS regression (assuming no strategic group level membership or effect), firms would seem to be best served by maintaining as little slack financial resources as possible (since the coefficient for current ratio is negative). Using HLM with grand mean centering, however, it seems that controlling for group and industry membership there is no substantive effect of slack financial resources on performance. The third option, using HLM and specifying group mean centering, suggests that firms should have more slack resources than other firms in their same strategic group to maximize their performance potential.

**Illustration Integrating Firm and Group Level Influences on Performance**

The tests of different model specifications when examining firm resources demonstrate that empirical assumptions can have dramatic effects on interpretations about the influences of firm resources on performance. Thus, researchers integrating firm and strategic group levels of analysis should exercise caution when implementing empirical techniques such as RCM where the choice of empirical specification or variable scaling reflects deeply impeded assumptions. For example, tests using OLS regression assumptions are not suitable for samples including multiple industries, and especially samples including multiple strategic groups nested within multiple industries.

<table>
<thead>
<tr>
<th>Resource Measure</th>
<th>OLS Regression Assumptions</th>
<th>HLM with Group Mean Centering</th>
<th>HLM with Grand Mean Centering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Standard error</td>
<td>Coefficient</td>
</tr>
<tr>
<td>Capital intensity</td>
<td>−.24</td>
<td>.14</td>
<td>−2.95</td>
</tr>
<tr>
<td>Patents</td>
<td>.39*</td>
<td>.16</td>
<td>.27</td>
</tr>
<tr>
<td>Current ratio</td>
<td>−1.54**</td>
<td>.29</td>
<td>1.28*</td>
</tr>
</tbody>
</table>

*\( p < .05 \).
**\( p < .01 \).
Both grand and group mean centering may be useful for research integrating the firm and group levels of analysis, but researchers should be cautioned that each specification corresponds to different theoretical assumptions regarding the influence of the group level of analysis. In this section, I illustrate the appropriate use of each centering specification in more complex models involving independent variables at multiple levels of analysis.

I should begin by noting that the individual level variables used to cluster firms may also be of interest as effects at the firm (as well as group) level of analysis. For example, Lawless et al. (1989) were the first to empirically test this idea, which they executed using separate regression and ANOVA tests to examine the efficacy of firm level influences on firm performance (using regression) and by testing the influence of independent variables that were clustered at the group level of analysis (using ANOVA). The use of RCM allows for a more integrative test of these levels of analysis that provides the ability to assess how group level characteristics may affect relationships at the firm level of analysis using a single analytical technique.

Grand mean centering is appropriate if the research goal is to simply control for the effects at the firm level of analysis. For example, R&D intensity has also been conceptualized as an important firm-level variable (e.g., Wu, Levitas, & Priem, 2005), and the theoretically defined groups used to illustrate RCM/HLM are based on the importance of R&D intensity at the group level (Ketchen et al., 1993; Bantel, 1998). The effect of R&D intensity at the group level, above and beyond the effects at the individual level, can be assessed using RCM/HLM. This could reveal whether strategic group characteristics explain variance in performance beyond that explained by firm characteristics. If this is the primary research goal, R&D intensity at the firm level of analysis should be modeled using grand mean centering. Thus, the following model is estimated:

\[
\text{Performance}_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{R&D intensity}_{\text{grand mean centered}})_{ijk} + e_{ijk}
\]

\[
\pi_{0jk} = \beta_{00j} + \beta_{01}(\text{group average R&D intensity})_j + r_{0ij}
\]

\[
\pi_{1jk} = \beta_{10j} + r_{1ij}
\]

\[
\beta_{00k} = \gamma_{000} + U_{00k}
\]

Table 6 displays the results for the tests of group main effects controlling for firm level effects. As shown in this table, group R&D intensity had a
significant effect on performance above and beyond firm level R&D intensity. Specifically, lower group level R&D intensity was associated with higher firm level ROA.

Grand mean centering is appropriate if the research goal is to simply control for the effects of firm level resources (as illustrated in the previous example), but researchers should be cautioned that grand mean centering models produce level-1 slopes that are actually composites of the within-group and between-group relationships among the independent and dependent variable (Hofmann et al., 2000). This inclusion of both the within-group relationship as well as the between-group relationship in the level-1 slope estimates can result in spurious cross-level interactions (Hofmann & Gavin, 1998). Thus, researchers interested in testing the moderating effects of strategic group characteristics on firm level relationships should rely on group-mean centering. This decision necessitates a number of empirical nuances that I illustrate in the following example.

I include R&D intensity at the firm level of analysis using group mean centering. Next, I test the influence of a group level variable (strategic group size) as a main (significant predictor of intercept) and moderating (significant predictor of slope) effects. Size of a strategic group has been an important group level characteristic historically and theoretically (Mas-Ruiz, Nicolau-Gonzalbez, & Ruiz-Moreno, 2005). To avoid the potential for spurious effects, it is necessary to introduce group mean R&D intensity back into the level-2 equations (Hofmann et al., 2000). Thus, the following series of equations is estimated:

$$\text{Performance}_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{R&D intensity}_{\text{group centered}})_{ijk} + e_{ijk}$$

$$\pi_{0jk} = \beta_{00j} + \beta_{01}(\text{group average R&D intensity})_{j} + \beta_{02}(\text{group size})_{j} + r_{0ij}$$

$$\pi_{1jk} = \beta_{10j} + \beta_{11}(\text{group size})_{j} + r_{1ij}$$

Table 6. HLM Results Integrating Firm and Strategic Group Levels Testing Group Main Effects with Grand Mean Centering at Level-1.

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average ROA, $\gamma_{000}$</td>
<td>-38.69</td>
<td>9.27</td>
<td>-4.18*</td>
</tr>
<tr>
<td>R&amp;D intensity (group level), $\gamma_{010}$</td>
<td>-8.16</td>
<td>2.27</td>
<td>-3.59*</td>
</tr>
<tr>
<td>R&amp;D intensity (firm level), $\gamma_{100}$</td>
<td>-25.61</td>
<td>7.60</td>
<td>-3.37*</td>
</tr>
</tbody>
</table>

*p < .01.
Table 7. HLM Results Integrating Firm and Strategic Group Levels Testing Group Moderation Effects with Group Mean Centering at Level-1.

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average ROA, $\gamma_{000}$</td>
<td>-8.39</td>
<td>3.93</td>
<td>-2.14</td>
</tr>
<tr>
<td>Group size (main effect), $\gamma_{010}$</td>
<td>-0.08</td>
<td>0.07</td>
<td>-1.19</td>
</tr>
<tr>
<td>R&amp;D intensity (group level), $\gamma_{020}$</td>
<td>-8.19</td>
<td>1.56</td>
<td>-5.26**</td>
</tr>
<tr>
<td>R&amp;D intensity (firm level), $\gamma_{100}$</td>
<td>3.29</td>
<td>1.39</td>
<td>2.36*</td>
</tr>
<tr>
<td>Group size (moderating effect), $\gamma_{110}$</td>
<td>-0.86</td>
<td>0.18</td>
<td>-4.72**</td>
</tr>
</tbody>
</table>

* $p < .05$.  
** $p < .01$.  

\[
\beta_{00k} = \gamma_{000} + U_{00k} \\
\beta_{01k} = \gamma_{010} + U_{01k} \\
\beta_{02k} = \gamma_{020} + U_{02k} \\
\beta_{10k} = \gamma_{100} + U_{10k} \\
\beta_{11k} = \gamma_{110} + U_{11k}
\]

Table 7 displays the results for the tests of group level moderation. As shown in this table, group size was a significant moderator of the relationship between firm level R&D intensity and ROA. Specifically, the relationship between firm R&D intensity and ROA was stronger in smaller groups than in larger groups. This illustration highlights a number of assumptions and empirical specifications that researchers should be cognizant of when testing many of the possibilities suggested in Table 2.

DISCUSSION

These illustrative tests suggest that RCM/HLM provides a powerful tool toward further integration of the firm and group levels, but considerable attention to detail conceptually and empirically is needed to fully utilize this technique. Researchers must clearly articulate their conceptualization of how firms and strategic groups interact, and choose the appropriate
methodological approach and specification that matches their assumptions. Doing so will facilitate knowledge development about the determinants of performance, while failure to do so will impede research progress. Researchers have noted that strategic groups research could benefit from use of ‘state of the art’ methods bridging research streams (Ketchen, Snow, & Hoover, 2004), and the ideas presented in Table 2 provide a starting point for such analysis. Failure to incorporate advancements in statistical analysis such as RCM, however, is likely to impede progress about the influence of multilevel phenomena on firm performance.

Researchers who do choose to utilize RCM must take care to articulate their research question carefully and apply this statistical technique correctly. If the research question examines the effects of strategic group characteristics controlling for the effects of firm resource measures, then RCM applications such as HLM with grand mean centering provide an appropriate match between conceptual aim and empirical technique. If researchers are seeking to truly integrate firm resource measures and the moderating effects of strategic groups membership or the moderating effects of strategic group characteristics (e.g., group size), then RCM techniques such as HLM with group mean centering provide an appropriate statistical tool. By using group mean centering, future researchers also have an appropriate tool that may be useful to answer researchers’ call to examine the role of the strategic group as a competitive referent (Ketchen et al., 2004).

Future research would benefit from additional theorizing and advancements in the measurement of the strategic group level. I relied on deductively defined strategic groups to create theoretically viable clusters that would have meaning across a variety of industries. While this is an established framework built on considerable strategic management thought, this empirical strategy is still subject to criticisms that group level phenomena were measured via aggregates of individual firm characteristics. Future research should work to integrate other forms of measuring strategic groups that are able to avoid issues associated with aggregating firm level variables while still being theoretically viable and allowing for empirical research spanning multiple industries. For example, cognitively defined strategic groups may be one possible option for integration in future research spanning multiple levels of analysis. Under this method, manager’s cognitive maps are key to assigning group membership (e.g., Reger & Huff, 1993). In a study examining changes in the banking industry, Reger and Palmer (1996) surveyed 25 top managers in the banking industry and solicited their judgments as to competitors in four traditional industry segments (commercial banks, thrifts, brokerages, and credit unions).
While this approach avoids aggregation issues, applying this technique to a multiple industry setting would require an ambitious sample (replicating the 12 industries using the same number of executives as Reger and Palmer, 1996 would require participation from 300 senior executives). Perhaps one possible method to capture managers’ cognitions without direct participation is through computer-aided content analysis. For example, Osborne, Stubbart, and Ramaprasad (2001) used computer-aided content analysis of presidents’ letters to shareholders to identify themes that were then clustered into cognitive strategic groups. While this approach follows from inductive, rather than deductive logic, it does provide for possibilities of analyzing large numbers of firms that could then be integrated into multiple groups nested within multiple industries.

Future research could advance the findings of this study by testing firm, strategic group, and industry level variables in a single analysis. While I controlled for the contextual effects of industry membership by using three level HLM, future research could illustrate the main and moderating effects of industry level variables. For example, measures of resource abundance (i.e., munificence), volatility (i.e., dynamism), and complexity (cf. Dess & Beard, 1984; Keats & Hitt, 1988; Palmer & Wiseman, 1999) have been common in the strategy literature, and I believe that empirical research integrating such effects with strategic group characteristics would be a valuable contribution to the literature.

Finally, recent applications of RCM to the strategic management literature have demonstrated the use of this technique to more appropriately model firm performance over time (e.g., Misangyi et al., 2006; Short et al., 2006). The group level of analysis, however, has not been taken into account to date. Testing the degree to which group characteristics influence firm performance over time would provide a contribution to both the variance components and strategic groups literatures.

CONCLUSION

The RBV of the firm and strategic groups research are two key conceptual frameworks that have encouraged a wealth of empirical studies examining the determinants of firm performance. Unfortunately, there have been far fewer studies that have incorporated both views to provide a truly multilevel investigation. While advanced statistical techniques exist that would allow for the integration of these two levels, strategic management researchers have been slow to utilize such techniques. The suggestions outlined in this
chapter provide an empirical blueprint to allow researchers to appropriately test characteristics that incorporate these important levels of analysis. I also hope that my illustrations using RCM will encourage future researchers to utilize a sophisticated empirical technique for integrating research at multiple levels of analysis.

ACKNOWLEDGMENT

The author would like to thank Donald Bergh and Dave Ketchen for their thoughtful comments on previous drafts of this manuscript.

REFERENCES


SPECIAL SECTION ON METHODS IN INTERNATIONAL STRATEGY RESEARCH
MODELING INTERNATIONAL EXPANSION

Xavier Martin, Anand Swaminathan and Laszlo Tihanyi

ABSTRACT

Strategy deals with decisions about the scope of the firm and related choices about how to compete in various businesses. As such, research in strategy entails the analysis of discrete choices that may not be independent of each other. In this paper, we review the methodological implications of modeling such choices and propose conditional, nested, mixed logit, and hazard rate models as solutions to the issues that arise from non-independence among strategic choices. We describe applications with an emphasis on international strategy, an area where firms face a multiplicity of choices with respect to both location and mode of entry.

Strategy research addresses decisions about the allocation of resources for the purpose of developing and maintaining competitive advantage that enables superior corporate performance (Hofer & Schendel, 1978). It places emphasis on large, discrete moves such as the decision to enter (or withdraw from) a certain business, or the choice to undertake such investment via a given mode of governance. As a result, empirical samples may be obtained by observing any number of investments undertaken by a set of firms over time. However, for theoretical as well as measurement reasons, most
empirical studies focus on a subset of possible choices and associated predictors. This implies the analysis of a partial set of discrete choices while ignoring other choices when the included and omitted choices may in fact not be independent of each other.

For instance, research on product–market diversification may examine the behavior of firms diversifying from one industry into another. In that case the sample may consist of firms that were initially present in a given industry or set of industries, and that diversify into a given industry or set of industries. The means of diversifying may vary as well. Product–market diversification may occur organically, via acquisition, and/or via alliances and joint ventures with other firms. Yet many studies focus exclusively on a given modality, such as acquisitions.

This diversity of choices is especially prominent for research in international strategy, which deals with the antecedents and consequences of decisions about where to operate internationally, and how to govern such operations. Firms that consider expanding internationally, often do so to respond to international competition. Firms may implement their international expansion by choosing among a wide array of potential host countries. Indeed, the range of location opportunities has increased in recent decades as the unraveling of the Cold War and developments in physical and informational infrastructure have opened new nations to foreign trade and investment. For the same reasons, foreign investors make increasingly specific decisions not only about countries of their operations; but also about specific locations (region, city, or neighborhood) within selected countries (Chang & Park, 2005; Shaver & Flyer, 2000).

International strategy also entails a diverse set of choices about the governance of activities abroad, which are commonly referred to as mode of entry (MOE) decisions. MOE choices include the wholly owned subsidiary; collaboration with a partner as in an alliance or equity joint venture; contractual arrangements such as licensing or franchising; and trade solutions such as exporting or importing. Furthermore, wholly owned as well as partially owned subsidiaries may come about as a result of either the acquisition of an existing business, or the creation of a new local organization where none existed before (often labeled, accordingly, greenfield entry). In this paper, we first review typical research designs that have been used to model international expansion. Second, we describe international expansion as a firm-level choice process and suggest appropriate modeling strategies. Finally, we discuss implementation issues in modeling international expansion.
CURRENT RESEARCH DESIGNS FOR STUDYING INTERNATIONAL EXPANSION

Empirical research on international expansion typically focuses on a subset of countries. In some instances, the analysis is limited to a single home country (or region) and a single host country (or region). Such studies may be motivated by the salience of this pair of locations, where salience may result from the presence of critical industry participants or specific competitive or regulatory conditions. For instance, Martin, Swaminathan, and Mitchell (1998) examined the propensity of Japanese automotive suppliers to build plants in the United States and Canada (treated as a common host location by virtue of a long-standing free trade pact in automobiles) – thus capturing learning and competitive effects within a subset of the worldwide automobile industry that was the largest in population and sales, and where distinctive practices in buyer–supplier relationship management existed. In other instances, authors study investment from multiple home countries into a single host country. Such studies often rely on databases collected by governments for the purpose of tracking inbound investment, including the U.S. International Trade Administration’s report on “Foreign Direct Investment in the United States, (various years) Annual Transactions” (Shaver, 1998; Chung & Alcácer, 2002). Yet others study investments from a single home country into multiple, heterogeneous host countries. These studies are often based on industry sources such as corporate directories or on national financial databases. For instance the Toyo Keizai Directory has been used in a number of recent studies on the foreign direct investment behavior of Japanese firms (Hennart & Reddy, 1997; Isobe, Makino, & Montgomery, 2000; Delios & Beamish, 2001; Berry, 2006). Between them, these three categories of studies (single home to single host, multiple homes to single host, and single home to multiple hosts) far exceed in number those studies that examine both multiple home countries and multiple host countries – though the latter type of studies can be found in the study of industries with a predominantly global trade and investment profile, such as semiconductors (Martin & Salomon, 2003a).

The research designs described above may lend themselves to the study of causal mechanisms for firm-level outcomes at various levels. Independent variables that signify these causal mechanisms follow from different theories. Studies of multiple host countries, in particular, draw on macroeconomic predictors. The relative wages and other input factor costs
of a country, for instance, is a likely predictor of how attractive producing in that country may be. While it is much less common to use macroeconomic variables to describe home-country characteristics, other variables have also been used to predict outward investment. For example, the concentration of the industry in a firm’s home country has been used as an indicator of oligopolistic pressure to emulate rivals’ moves (Knickerbocker, 1973). The most common category of predictor variables, however, consists of firm-level characteristics. Two types of firm-level characteristics are especially prominent in research on international expansion (e.g. Martin & Salomon, 2003a): intangible assets, since internalization theory (a powerful theory developed by Buckley and Casson (1976) which uses arguments related to transaction cost economics and agency theory) predicts that foreign direct investment is justified when a firm possesses intangible assets that cannot be safely transacted across firm boundaries; and organizational learning through experience (Barkema & Vermeulen, 1998).

**SINGLE- AND MULTIPLE-MODE OF ENTRY STUDIES**

Likewise, different studies may sample different MOEs. Indeed, studies that focus on a single entry mode are the most common. For example, there has been a voluminous stream of research on joint ventures and other forms of alliances (e.g. Barkema, Shenkar, Vermeulen, & Bell, 1997; Contractor & Lorange, 1988, 2002). Other studies focus on a mix of wholly and partly owned subsidiaries, without making a particular distinction between these MOEs (e.g. Shaver & Flyer, 2000). Yet others study acquisitions alone (e.g. Markides & Ittner, 1994). However, the underlying theory largely emphasizes the comparative governance and strategic features of various MOEs. Wholly owned, joint venture and contractual modes can be arrayed along a continuum that trades off organizing and bureaucratic costs against market hazards (Dunning, 1995; Hennart, 1993; Williamson, 1991). As such, factors that encourage one MOE may encourage or discourage other MOEs, and a comprehensive comparison among these solutions is required to make sense of a given choice.

As with location decisions, MOE theory suggests a mix of firm- and country-level considerations. For instance, in internalization theory the ability to successfully operate and compete abroad is predicated on the possession of distinctive intangible assets – a firm-level feature. Whether
these assets are transferred to the firm’s own subsidiary or leased out via licensing will depend on the ability of the firm(s) involved to transfer the underlying knowledge (Kogut & Zander, 1993; Martin & Salomon, 2003b); but also on the efficiency and reliability of the intellectual property regime in the host country (Oxley, 1999). Thus, various MOE choices share explanatory variables.

A further complication arises from the distinction between acquisition and greenfield modes. Theoretically, these choices are substitutes in that both allow the firm to obtain control over assets in a foreign location. However, each offers distinct advantages. For instance, an acquisition may also be the means for a foreign investor to obtain local resources such as a brand or production expertise – but may be more costly.

Furthermore, while the means of obtaining control over a foreign operation (greenfield vs. acquisition) is theoretically separate from the level of ownership or control the firm seeks (wholly owned subsidiary vs. equity joint ventures), studies aggregate these MOE combinations differently. Some studies treat acquisitions as a separate choice than (greenfield) wholly owned subsidiary – thus ignoring whether the foreign investor acquires the whole or a portion of the target firm’s equity (e.g. Anand & Delios, 2002; Hennart & Park, 1993; Barkema & Vermeulen, 1998). Others aggregate acquisitions with wholly owned subsidiaries, comparing them with joint ventures – thus assuming that all acquisitions are of the whole of the target’s equity (Hennart & Reddy, 1997; Brouthers, Brouthers, & Werner, 2003). Relatively few studies separate greenfield from acquisition as well as wholly owned from joint venture choices (e.g. Kogut & Singh, 1988; Gatignon & Anderson, 1988).

It should be evident from our brief review that choices about international expansion are influenced both by firm- and country-level factors. Moreover these choices are made not individually, but from a choice set. Therefore their examination requires the use of statistical methods that take interdependence across choices into account.

MODELING INTERNATIONAL EXPANSION

We suggest appropriate models to simultaneously take into account firm- and country-level predictors of international expansion. We first introduce the conditional logit model and describe its uses and limitations. Second, we discuss the nested and mixed logit models that overcome many of the deficiencies of the conditional logit model. We also describe hazard rate
models that allow for multiple events in the same time period but do not allow for unobserved taste variations among firms. Third, we briefly address the choice of MOE. Finally, we discuss implementation of these models in three widely used statistical software packages, NLOGIT, SAS, and STATA.

The Conditional Logit Model

International expansion can be viewed as a firm-level choice process where both the characteristics of the chooser (the firm) and the choices (the countries in the choice set) influence the probability of a specific choice being made. The model suited to such choice problems is McFadden’s (1974) conditional logit model. In the case of international expansion, the probability that a firm \( i \) will enter country \( j \) can be depicted as:

\[
P_{ij} = \frac{\exp(\beta'x_{ij})}{\sum_{k=1}^{m} \exp(\beta'x_{ik})}
\]

(1)

where \( m \) equals the number of possible countries in the choice set and \( x_{ik} \) is a vector of observed variables relating to country \( k \). Eq. (1) has several desirable properties. First, the value of \( P_{ij} \) is always between 0 and 1. Second, the choice probabilities for entry into all countries sum to 1. The conditional logit model, however, suffers from at least three serious limitations. First, the model is incapable of representing random taste variation related to properties of the choosers, firms in this case. Thus, it is not optimal in analyzing or controlling for the effects of corporate heterogeneity, even though such heterogeneity is theoretically salient in (international) strategy research. Second, when comparing between any two alternatives say \( a \) and \( b \), in this case target countries for expansion, the model assumes that the ratio of the probabilities of choosing among \( a \) and \( b \) is not dependent on the attributes of any other alternative \( c \). This property of the conditional logit model, called the independence from irrelevant alternatives (IIA), is a questionable assumption when applied to international expansion. Third, the model does not accommodate cases where unobserved factors are correlated over time. This is a particularly important concern in modeling international expansion, which is best represented as a process that unfolds within each firm over time and one that is subject to firm-level learning effects.
The Nested Logit Model

Nested logit models can be used to address modeling issues raised by the IIA property of conditional logit models. For instance, the world can be divided into regions (based on geography, state of economic development, and so on) within which we can reasonably assume that the IIA property holds. Nested logit models can then be used to estimate the probability of choosing a nest (a region) and the alternatives (countries) within a nest. However, the first and third limitations above remain. A more realistic representation of the international expansion process involves the use of mixed logit models (McFadden & Train, 2000; Train, 2003) which allow us to address all three limitations of the conditional logit model.

The Mixed Logit Model

Mixed logit models allow for random taste variation, allow any pattern of substitution among choices, and can handle panel data with temporally correlated errors. Mixed logit probabilities can be expressed as follows:

\[ P_{ni} = \int L_{ni}(\beta) f(\beta) d\beta \quad (2) \]

where \( L_{ni}(\beta) \) is the logit probability evaluated at parameters \( \beta \).

\[ L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}} \quad (3) \]

and \( f(\beta) \) is a density function. \( V_{ni}(\beta) \) is the observed portion of the utility which depends on the parameters \( \beta \). If utility is linear in \( \beta \), then \( V_{ni}(\beta) = \beta'x_{ni} \). Then the mixed logit probability takes the form:

\[ P_{ni} = \int \left( \frac{e^{\beta'x_{ni}}}{\sum_{j} e^{\beta'x_{nj}}} \right) f(\beta) d\beta \quad (4) \]

Eq. (4) is equivalent to a weighted average of the logit formula evaluated at different values of \( \beta \) with the weights given by the density \( f(\beta) \) which can be specified as continuous or discrete. The mixed logit model can be derived from utility maximizing behavior in several ways. Below we describe the
random coefficient or random parameter logit model that has been used in recent applications to choice processes (Erdem, 1996; Revelt & Train, 1998; Chung & Alcácer, 2002).

**Random Coefficients Logit Model**

Using the random coefficients logit model the utility derived by a firm \( n \) from entering country \( j \) is given by:

\[
U_{nj} = \beta_n x_{nj} + \epsilon_{nj}
\]

where \( x_{nj} \) are observed variables that are associated with the country and the firm, \( \beta_n \) is a vector of coefficients of these variables for firm \( n \) representing that firm’s tastes, and \( \epsilon_{nj} \) is a random term that is independent and identically distributed (i.i.d.) extreme value. The coefficients are assumed to vary over firms in the population with density \( f(\beta) \). The firm \( n \) knows the value of its own \( \beta_n \) and \( \epsilon_{nj} \)'s for all countries \( j \) and chooses alternative \( i \) if and only if \( U_{ni} > U_{nj} \forall j \neq i \). The researcher observes the \( x_{nj} \)'s but not the \( \beta_n \)'s or the \( \epsilon_{nj} \)'s. The unconditional choice probability is therefore the integral of \( L_{ni}(\beta_n) \) over all possible variables of \( \beta_n \):

\[
P_{ni} = \int \left( \frac{e^{\beta_n x_{ni}}}{\sum_j e^{\beta_n x_{nj}}} \right) f(\beta) d\beta
\]

which is the mixed logit probability given by eq. (4).

The researcher has to specify a distribution \( f(\beta) \) for the coefficients and estimates the parameters of that distribution. In most cases (e.g. Revelt & Train, 1998), \( f(\beta) \) has been specified as normal or log-normal: \( \beta \sim N(b, W) \) or \( \ln \beta \sim N(b, W) \) with parameters \( b \) and \( W \) to be estimated. The log-normal distribution is useful if the mean effect \( b \) of any variable is known to have the same directional effect for all firms; by contrast, the normal distribution allows the mean effect to be either positive or negative depending on the firm. Exact maximum likelihood estimation is not possible since the integral in eq. (4) cannot be calculated analytically. Instead researchers have approximated the probability through simulation and maximized the simulated log-likelihood function (Erdem, 1996; Revelt & Train, 1998; Bhat, 2000).

The random coefficients logit model can easily be generalized to allow for the use of panel data and repeated choices regarding international expansion by each firm. The probability of international expansion by a firm is the
product of logit formulas, one for each time period. Lagged dependent variables can be added to the model without changing the estimation procedure. Thus one can estimate the extent of repetitive momentum in international expansion by a firm. If the choices of firms and other data are not observed from the beginning of the process, the researcher has to somehow represent the probability of the first observed choice, which depends on the previous unobserved choices. Heckman and Singer’s (1986) correction for sample selection bias can be used for this purpose.

**Hazard Rate Models**

Hazard rate models have been used to estimate the rate of expansion into a specific destination country (e.g. Martin et al., 1998). The instantaneous rate of expansion into a foreign country is given by:

$$ r_{ijk}(t) = \lim_{t' \to t} \Pr(t \leq t' | T \geq t) $$

(6)

The hazard rate $r_{ijk}(t)$ can be interpreted as the propensity for a firm $i$ to expand from home country $j$ to a host country $k$, at time $t$, given that it has not expanded to country $k$ before time $t$. The rate can be specified as a function of duration $t$ and a vector of firm-specific variables $x_i(t)$, a vector of home country variables $y_j(t)$, and a vector of host country variables, $z_k(t)$. The relationship between the rate $r_{ijk}(t)$ and the vectors $x_i(t)$, $y_j(t)$, and $z_k(t)$ is assumed to be loglinear to ensure non-negative rates.

$$ r_{ijk}(t) = f(t, x_i(t), y_j(t), z_k(t)) $$

(7)

The choice set in international expansion is made up of multiple countries or destination states. Usually multiple destination states are modeled as competing risks, but we typically do not have theories about how expansion into a particular country differs from expansion into another. Instead, we have theories about how the underlying characteristics of various locations (countries) influence the likelihood of expansion. One way to model international expansion within a hazard rate framework is to treat the firm-country pair as the unit of analysis in each time period. In this case, firms will have multiple records (as many as in their choice set) in each time period. Within each time period firm characteristics will be fixed but country characteristics will vary. Hazard rate models have an advantage over conditional, nested, and mixed logit models in that they allow for multiple expansion events for a firm in a given time period. They are also able to
incorporate observed variations in firm and country characteristics and unobserved heterogeneity among firms (Blossfeld & Rohwer, 2002). But they do not allow us to estimate unobserved variations in tastes among firms (the $W$ parameter in the random coefficients logit model described above).

Hazard rate models have been used extensively in the study of organizational change (Barnett & Carroll, 1995). International expansion is but a specific instance of corporate expansion, and more generally organizational change. We believe that much can be gained from recent developments in the study of organizational change (Barnett & Carroll, 1995). Organizations have been found to change in response to three different causal mechanisms: (1) organization-specific factors including their own previous history of change (Amburgey & Miner, 1992); (2) changes in the organizational environment (Romanelli & Tushman, 1994); and (3) the actions of other organizations, including competitors, in salient reference groups (Haveman, 1993; Martin et al., 1998). These influences can be combined in a heterogeneous diffusion model (Strang & Tuma, 1993) to study organizational change (e.g. Greve, 1995, 1996). The heterogeneous diffusion model is useful to depict organizational change in a population where behavior is contagious. The rate of change is modeled through four vectors, the intrinsic propensity to change, the susceptibility to contagion, the infectiousness of an organization that has already changed and proximity among organizations. International expansion offers unique benefits in studying organizational change because the state space (countries) is finite and clearly specified. This is not typically the case with other types of organizational change such as product–market diversification or technological change.

Data Quality and Statistical Software for Estimating Choice Models for Panel Data

International expansion is driven by firm-level factors including the firm’s history, its interactions with other firms in an industry, and location-specific factors. This is a process that unfolds over time and may involve repeated events for a particular firm in a particular country (e.g. Chang & Rosenzweig, 2001). The appropriate data to analyze this outcome is longitudinal data on the behavior of all firms within an industry and the characteristics of all potential host countries. Countries that do not experience any entry by foreign firms over the entire time period under consideration ought to be excluded from the choice set. The conditional,
nested and mixed logit models described above assume that a firm will make a single choice in any time period under consideration. Simultaneous expansion into multiple host countries is rare, but it does occur if we observe international expansion at sporadic time intervals. One solution to deal with such cases is to choose the time intervals such that in any single time interval, we observe that a firm experiences no more than a single expansion event. The other option is to employ hazard rate models that allow for multiple events by a firm in a particular time period. The statistical models described above have been implemented in at least three commonly used statistical software packages, *NLOGIT*, *SAS*, and *Stata*.

The *NLOGIT* package, an extended version of *LIMDEP* allows for the estimation of conditional logit, nested logit, and random coefficient logit models (Hensher, Rose, & Greene, 2005). There are limits to the number of choices that can be considered. The conditional logit model in *NLOGIT* considers a maximum of 50 alternatives. The nested logit tree allows for four levels, labeled trunks, limbs, branches, and alternatives, respectively. Nested logit models estimated by *NLOGIT* may have up to a maximum of 5 trunks, 10 limbs, 25 branches, and 100 alternatives. The random coefficients model allows one to consider up to 100 alternatives. \( f(\beta) \) can be specified as normal, log-normal, uniform, triangular, and non-stochastic (zero variance). The default number of draws for constructing the simulated maximum likelihood is 100 but this can be changed to a higher number. Train (1999) recommends several hundred random draws while Bhat (2001) recommends 1,000 random draws. Since such an exercise is computationally intensive, Greene (2002) recommends that the number of random draws be set below 20 for exploratory purposes and increased once a final model specification is selected. The program allows one to use Halton intelligent draws which reduces the number of draws required by as much as 90% (Bhat, 2001).

The *MDC* procedure within the *SAS* statistical software program also estimates conditional logit, nested logit, and random coefficient logit models. The nested logit specification allows for three levels. The random coefficient logit model allows the researcher to specify \( f(\beta) \) as a normal, log-normal, or uniform distribution. The default number of draws is 200 and Halton sequence generation is available. The number of choices is unrestricted in the MDC procedure.

*Stata* provides both conditional logit and nested logit models through the *clogit* and *nlogit* commands. The nested logit model allows for three levels within the choice tree. No restrictions are placed on the number of choices.

Hazard rate models that take into account panel data, time varying covariates, and repeatable events can be estimated using the *SURV*
procedure in *NLOGIT (LIMDEP)* and the *streg* procedure in *Stata*. Both procedures allow for a variety of parametric distributions for duration dependence in the hazard rate. The *LIFEREG* procedure in *SAS* estimates accelerated failure time models where the dependent variable is the duration to the event and not the hazard rate. The implementation in *SAS* does not allow for time varying covariates or repeatable events.

*Modes of Expansion*

The choice of the MOE can be modeled using the same methods as the choice of the country that a firm expands into. Theory should determine the appropriate modeling strategy. If the choice of destination country and the choice of the MOE are considered to be sequential choices, the choices should be modeled conditional on the other choice having been made first. For instance, if one expects that firms first choose the country to expand into, then the models of MOE choice should be estimated using data comprising the subset of countries into which a firm has expanded. It is also possible that firms have a characteristic MOE that they have honed through experience. In this case the choice of MOE ought to be modeled before the choice of countries that a firm expands into.

If the choice of destination country and the choice of the MOE are simultaneous, the same statistical models can be used with an extended choice set made up of all country–MOE combinations for each firm.

**INTERNATIONAL EXPANSION AND CORPORATE CHANGE: THEORY AND RESEARCH DESIGN**

To illustrate the benefits of the above models, we consider Dunning’s eclectic paradigm (Dunning, 1980, 1981, 1988). This framework constitutes the most systematic attempt to arrive at a comprehensive theory of international expansion. Dunning explained the specific patterns of a firm’s international expansion in terms of its ownership-specific advantages (O), the location-specific advantages (L) of the countries that it expanded into, and its internalization advantages (I) which imply that is in the interest of a firm that possesses ownership-specific advantages to transfer them across countries within its own organization rather than through exporting or licensing their use by a host-country firm. The OLI framework for explaining patterns of firm-level international expansion thus implicitly
assumes that the process is driven by observed variation in firm- and location-specific characteristics (Dunning, 1979). In other words, firms with varying firm-specific characteristics (O) choose among locations with varying location-specific (L) characteristics.

The conditional logit model (McFadden, 1974) has been designed to model such choices. Internalization (I) considerations are related to the MOE and can be easily accommodated within the conditional logit model by characterizing the choice as one of the location combined with the MOE. For instance, if the three modes of entry being considered are wholly owned, joint venture, and licensing and the number of countries that the firm could expand into is \( N \), then the choice set is made up of \( 3N \) elements. Mixed logit models offer the added benefit of allowing for unobserved taste variations on the part of firms. The use of these choice models would allow the researcher to disentangle the effects of O, L, and I factors on firm-level international expansion and to determine their relative influence.

The use of hazard rate models that incorporate heterogeneous diffusion processes will allow us to develop and test additional predictions within the OLI framework. In the case of international expansion, O, L, and I variables will likely influence the intrinsic propensity to change. Proximity effects can be used to examine whether firms from the same home country exhibit similar patterns of international expansion. Oligopolistic reaction, mimetic and bandwagon theories would suggest that firms from the same industry would exhibit strong proximity effects. Firm-specific characteristics such as size, level of intangible assets, or performance may also influence the infectiousness of a firm’s international expansion decisions and its susceptibility to the actions of others (e.g. Haunschild & Miner, 1997). We believe that modeling international expansion as a diffusion process offers opportunities to develop and test new theoretical predictions about international expansion.

In summary, we have shown that the study of corporate expansion, and specifically international expansion, involves the modeling of choice sets that entail considerations at multiple levels (schematically, country and firm effects). Modeling approaches that ignore the interdependence among sets of choices (such as countries and modes of expansion) risk misspecifying these interdependencies. We described four methods – conditional logit, nested logit, mixed logit (random parameter model), and hazard rate – and discussed the extent to which they address three key assumptions required to suitably model interdependent choices. We also identified statistical software packages that have implemented these methods. Finally, we note that attention to the statistical modeling of international expansion.
contributes not only to the testing of comprehensive theories of (international) corporate expansion, but also to the extension of the underlying theories themselves.

ACKNOWLEDGMENT

We thank Chih-Ling Tsai for his comments on an earlier draft of this paper.

NOTES

1. As in other (non-international) strategy research, studies also vary in the range of industries from which they sample firms. Single-industry or single-sector studies are quite common. Some specialized MOEs, including turnkey operations and management contracts, are only applicable to some industries.

2. The logic underlying this particular prediction is ambiguous, and as such it has been largely abandoned in recent literature. Nevertheless, this home-country measure is still occasionally used, mostly as a control variable.

REFERENCES


STATIC TRIANGULAR SIMULATION AS A METHODOLOGY FOR INTERNATIONAL STRATEGIC MANAGEMENT RESEARCH

George S. Yip, G. Tomas M. Hult and Audrey J. M. Bink

ABSTRACT

Emerging thoughts and models in strategic management increasingly involve complex hypotheses at different levels of analysis and multiple sides of relationships. Such complexities often result in less than ideal empirical testing, with the ensuing implications being limited or sometimes even wrong. One such case is global relationship management (GRM). The effective implementation of GRM has been argued to be a principal source of a firm’s value creation but the testing of GRM scenarios have been very limited. Using GRM as a case example, we introduce a new methodology to the strategic management literature that alleviates many of the limitations of existing techniques – static triangulation simulation (STS). A series of GRM hypotheses are briefly introduced and then tested via the STS technique. Starting values for the simulation, based on input from companies, are included from two sides of each GRM relationship.
(customer and supplier) and two levels (company and account) from each side. Such elaborate testing is typically not feasible via “normal” methodology – the STS technique, however, allows for a robust assessment of the different drivers that affect GRM outcomes.

INTRODUCTION

Global relationship management (GRM) has become an increasingly important tool for companies in their quest to achieve superior market success. And it is not simply an extension of key account management to the international marketplace. An intimate understanding of critical industry, strategy, and organizational factors underlie the operations of value-added GRM. However, empirically testing the critical relationships involving industry, strategy, and organizational factors across management levels within organizations (multiple levels) and between organizations (multiple sides) creates methodological difficulties. Specifically, obtaining multiple responses from the same organization is a tough task in many research projects. Obtaining multiple responses at different levels of hierarchy in an organization that are matched with the same multiple levels between organizations at a large scale is almost impossible. As such, researchers opt to either ignore critical research questions pertaining to scenarios such as these (i.e., multiple levels and multiple sides) or they opt to test only some of the relationships where data can be “easily” obtained. This renders the model examined incomplete (underspecified), the results weak, and the implications potentially flawed. What is the methodological solution?

We answer this question by introducing a static triangular simulation (STS) technique to the strategic management literature and illustrate its use by testing what drives the outcomes of GRM relationships. To that end, a brief overview of GRM is appropriate to set the stage for the use of the STS technique. Within the GRM framework, global customers and suppliers can be defined in a number of ways, with the caveat that a GRM system places the focus at the business-to-business level, not at the level of the end-consumer. At its broadest, global customers and suppliers are defined as any customer or supplier representing a multinational corporation (MNC) that is buying or selling in multiple countries. However, only those MNCs that actually globally coordinate their purchasing or selling can be considered global customers or suppliers. Importantly, global customer or supplier management programs are often applied at the regional level (e.g., Europe, South America). Conceptually, the issues at the regional and global levels are
treated similarly in a GRM program. As such, we define global customer management (GCM) as an organizational architecture (organizational form and process) by which the worldwide activities of the MNC (which serve a given multinational customer) are coordinated centrally by one person (or team) within the supplying company. Similarly, we define global supplier management (GSM) as an organizational architecture (organizational form and process) by which the worldwide activities of the MNC (which serve a given multinational supplier) are coordinated centrally by one person (or team) within the buying company.

The literature about managing global customers has converged on the term “global account management (GAM)” (e.g., Arnold, Birkinshaw, & Toulan, 2001 and Birkinshaw, Toulan, & Arnold, 2001; Hennessey & Jeannet, 2003; Montgomery & Yip, 2000; Verra, 2003; Wilson, Speare, & Reese, 2002; Yip & Madsen, 1996) as the terminology in many, although not all, companies (Belz & Senn, 1999). In terms of managing global suppliers, both the literature and practitioners tend to talk about “global supply chain management” (e.g., Cannon & Homburg, 2001; Cohen, Fisher, & Jaikumar, 1989; Houlihan, 1986; Schary & Skjøtt-Larsen, 2001). Both GCM and GSM can be viewed as aspects of relationship management (e.g., Parvatiyar & Gruen, 2001; Yip & Madsen, 1996) and GCM can be viewed as an extension of relationship management as well (e.g., Grönroos, 1997; Kalwani & Narayandas, 1995). Hence, we prefer to use the term global relationship management to apply to the strategic management of either customers or suppliers. While the term GAM tends to refer to formal programs, in many companies GCM and GSM may be less formal than a specific program. Also, we view GAM as only one aspect, albeit the most important one, of all that is involved in managing a global relationship.

In this study we apply the STS technique by integrating the two broad streams of research on global sales management and global purchasing, and, in particular, to test for parallels in how MNCs successfully manage their GCM and GSM activities. More specifically to the STS technique, we seek to remedy drawbacks of previous studies, particularly that of not obtaining measures from (a) both sides of the account manager/managed account dyad and (b) both the account level and the company level. Hence, our study is three-dimensional, covering the buying side/selling side dimension, the manager/managed dimension, and the account level/company level dimension. The effective testing of the three-dimensional framework is accomplished via the STS technique. Thus, we offer two main contributions: (1) the introduction of the STS technique to strategic management research and (2) an assessment of what drives outcomes in customer and buyer GRM relationships.
**MODEL TO BE TESTED VIA STS**

As a setup to the use of the STS technique, we briefly develop the GRM model to be tested. For ease of understanding, the model is structured to be the same across the GCM and GSM settings. The testing entails two parallel models that each corresponds to the same type and form of testing but involves different entities in the supply chain. The GCM model is from the viewpoint of a global supplier seeking to manage its global customers, while the GSM model is from the viewpoint of a global customer seeking to manage its global suppliers. The linkages are depicted in Fig. 1.

---

**Notes**

1. The top left cell shows the Manager (MR1) entity at the upper (company or business unit) level.
2. MR1 has a direct relationship with the Managed (MD1) entity at the same upper (company or business unit) level.
3. MR1 has a direct vertical relationship with the Manager (MR2) at the lower account level.
4. Similarly, MD1 has a direct vertical relationship with the Managed (MD2) at the lower account level.
5. The most intense relationship (shown by the bold horizontal arrow) occurs between the Manager (MR2) and the Managed (MD2) at the lower, account level.
6. Weaker relationships occur when across both companies and levels, as shown by the two diagonal, dashed line arrows.
7. The notations below refer to the types of respondents for the six different questionnaires that we used in the study:

   - C1 = Global Customer Account Management – Company Level Director of Global Customer Accounts
   - C2 = Global Customer Account Management – Account Manager
   - C3 = Global Customer Account Management – Account Level Customer
   - S1 = Global Supplier Account Management – Company Level Director of Global Supplier Accounts
   - S2 = Global Supplier Account Management – Account Manager
   - S3 = Global Supplier Account Management – Account Level Supplier

---

*Fig. 1.* Matrix of GCM and GSM Relationships.
Analysis of Different Levels (across and within Companies)

GRM programs operate both across companies and at different levels within companies. Across companies, and inherent in their nature, GRM programs have the dyadic aspect of involving both the company that manages the account and the company whose account is being managed (i.e., a customer managing a supplier – GSM, or a supplier managing a customer – GCM). Within companies, there are at least two major levels of GRM involvement: company or business unit and the account manager/managed level. First, there is usually some overall design for GRM at the company or business unit level. We allow for the business unit as the top level in some cases, as in many large multidivisional companies the business unit can function as a virtually autonomous company (at least for the purposes of GRM). Also, there are often separate GRM programs for different divisions, which fits our argument below that the industry of the supplier is a key determinant of the nature of GRM programs. Hence, business units in the same company but in different industries are likely to design different GRM programs.

Second, the key GRM relationship is at the account level, typically represented on the manager side by an account management team, reporting up to the company or divisional level; and on the managed side by a selling or buying team. We recognize, of course, the multi-level, multi-functional nature of decision-making units (DMU) (whether buying or selling), but most DMUs have a focal or lead manager, usually at a level below the head of the company or division. Hence, the GRM relationship can be represented in the $2 \times 2$ matrix shown in Fig. 1. The most intense relationship is shown by the thick double-headed arrow between MR2 and MD2, the account manager and the account being managed. The other primary relationships are likely to be those shown by the other solid arrows, while the typically less frequent relationships (across organizations and levels) are show by the dashed arrows.

Third, geography constitutes another dimensional level, with the key distinction being that between the headquarters (HQ) country and that of subsidiaries. Typically, a global customer has its global buying team, if there is one, located in the country of its HQ, although some divisional HQs may increasingly not be in the same country as the company HQ. Similarly, a global supplier usually locates at least some members of its selling team in the customer’s HQ country, although with strong relationships to its own HQ country. For example, Hewlett-Packard, as a supplier, locates its global account managers in the HQ country of its global customer accounts, but
these HP GAMs are strongly supported at HP’s HQ in the United States by Headquarters Account Managers, “HAMs” (Yip & Madsen, 1996; Yip, 2003, p. 167). There are exceptions. Standard Chartered Bank, a major U.K.-based multinational, locates many of its global account managers for European customers in Singapore, for cost reasons and because of the company’s strong Asian heritage (the bank having been founded in Hong Kong). Another aspect of geography is that GRM relationships are implemented in both HQ and subsidiary countries. Also, there is often, even usually, resistance from subsidiaries to the centralizing effects of most GRM programs.

Relationships Examined via the STS Technique

Both direct effects and strategic fit effects are tested via the STS technique to analyze relationships inherent in a GRM relationship. These 15 relationships are grouped into 10 hypotheses. Each hypothesis is listed below and more fully developed in Appendix 2.

Supplier Industry (H1a): Stronger globalization drivers in a supplier’s industry result in better marketing outcomes from a GCM program.
Customer Industry (H1b): Stronger globalization drivers in a customer’s industry result in better marketing outcomes from a GSM program.
Demand for GRM Services (H2a): Greater demand by customers for GCM services results in better marketing outcomes from a GCM program.
Customer Demand (H2b): Greater demand of customers with a GSM program for GRM services from their suppliers results in better marketing outcomes from a GSM program.
Use of GRM Program by Supplier (H3a): More extensive use by a supplier of aspects of a GCM program results in better marketing outcomes from a GCM program.
Use of GRM Program by Customer (H3b): More extensive use by a customer of aspects of a GSM program results in better marketing outcomes from a GSM program.
Extent of Supplier Business (H4a): More business for a supplier with global customers results in better marketing outcomes from a GCM program.
Extent of Customer Business (H4b): More business for a customer with global suppliers results in better marketing outcomes from a GSM program.
Supplier Power (H5a): Greater relative power for a supplier results in better marketing outcomes from a GCM program.
Customer Power (H5b): Greater relative power for a customer results in better marketing outcomes from a GSM program.

Fit Variables

Global Strategy Fit (H6): Closer fit between a supplier’s and a customer’s company global strategies results in better marketing outcomes from either’s GRM program.
Global Organization Fit (H7): Closer fit between a supplier’s and a customer’s company global organization results in better marketing outcomes from either’s GRM program.
Demand Fit (H8): A closer fit between a customer’s demand for GRM services and a supplier’s perception of that demand results in better marketing outcomes from either’s GRM program.
Program Match (H9): A closer fit between a supplier’s and its customer’s GRM programs results in more positive performance improvements from either’s GRM program.
Program Customization (H10): Greater customization of the company level GRM program, for the individual account level, for either suppliers or customers, results in better marketing outcomes for either’s GRM program.

DESCRIPTION OF THE RESEARCH DESIGN

The complexity of the relationships in H1 through H10, albeit supported theoretically, calls for a unique research design. Traditional primary (survey) or secondary (archival) data cannot address each aspect of what is hypothesized. The result is that researchers often use proxies for some or all of the variables or chose to test a portion of the framework (often from one side of the dyad and at one level of analysis). Such empirical assessment, at best, renders incomplete findings and, at worst, leads to flawed implications. As an example of a methodology to offset some of the limitations with current primary and secondary research methods, we use STS as an example of a technique that can overcome many of the limitations of past strategic management research on supply chains.

Specifically, our model called for a research design that accommodated two sides (customer and supplier), two levels (company and account), and
two types of programs (GCM and GSM) – resulting in a “2-cubed” design. Furthermore, extensive data were needed for each of the 2-cubed entities. No previous studies had collected data from all six entities. Indeed, many collected data from just one entity, e.g., the company-level supplier view of a GCM program (Montgomery & Yip, 2000) and from the global account manager view of a GCM program (Arnold et al., 2001; Birkinshaw et al., 2001). These prior studies gained large (100+) sample sizes at the expense of comprehensiveness in coverage of entities.

In addition, it is now well established that more than one key informant is needed to develop reliable measures (Baumgartner & Steenkamp, 2006; Moriarty & Bateson, 1982). For example, in a study of corporate culture, customer orientation, and innovativeness, Deshpandé, Farley, and Webster (1993) used a “quadrad” design of double dyads of customers and suppliers. However, they used the same level of customer or supplier executives and hence covered only one dimension, while our research design covers three dimensions. Specifically, our study methodology has three components: (1) collected both qualitative and quantitative information in the 2-cubed design, (2) applied “static triangular simulation” to simulate the potential values within the variance established by the surveys in the GCM and GSM relationships, and (3) used fit-based moderator hierarchical regression analysis to empirically test the hypotheses.

Unit of Analysis

To gain access to both sides of a customer–supplier dyad is difficult, especially when asking a supplier to provide access to its customers. Even multi-level access within the same company has its difficulties, whether the initial contact is at the company level or the account manager level. Hence, we decided to use a case-based method, which allowed us to build a relationship with each company. This relationship then gave us access to multiple account managers within the company, and allowed us to nurture account managers’ willingness to provide further access across the dyad to multiple customers or suppliers.

Specifically, we sought to recruit respondents (from several companies) in the MR1 cell in Fig. 1, the company level account manager, typically the head of GCM or the head of GSM. Then we asked each company’s GRM head to identify up to four individual global accounts that would participate in our study as well as to introduce us to the individual global account managers. In turn, we asked these individual global account managers to
introduce us to their corresponding global customer or supplier on the other side of the dyad. To collect the data we conducted both semi-structured interviews and administered a multi-part questionnaire. As illustrated in Fig. 1, we interviewed and surveyed six types of respondents: C1 – the company level director of global customer accounts, S1 – the company level director of global supplier accounts, C2 – the global customer account manager, S2 – the global supplier account manager, C3 – the customer being managed at the global account level, and S3 – the supplier being managed at the global account level. Ideally, we would have also gained access to respondents in the MD1 cell, the company level head of global customer or supplier management on the other side of the dyad, but we soon discovered that seeking such access overly strained our relationship with the primary contacts.

Companies Researched

We developed a list of companies to approach for participation, based on our researched knowledge of the likelihood that they would have significant GCM or GSM programs. We succeeded in gaining cooperation from nine major multinational companies, all of them headquartered in Europe, except for the European division of one U.S. Company, Xerox. Table 1 lists the five global suppliers and the four global customers. Table 1 also lists the 18 global customers of the 5 global suppliers and the 9 global suppliers of the 4 global customers. Several of these individual customers and suppliers are U.S. companies, although the majority of the companies are European. Interestingly, some companies participated as both customers and suppliers. For example, at WPP we were able to research its relationship with Vodafone as both a customer of WPP and a supplier to WPP. Less directly, we researched Unilever as a global supplier and met this company again as a global customer of two other global suppliers.

USING THE STATIC TRIANGULAR SIMULATION TECHNIQUE

Data Used as Input into the Simulation

The STS technique has been used previously in studies on supply chains, albeit not as rigorously and in such a large scale as our case example (e.g.,
Mollenkopf, Closs, Twede, Lee, & Burgess, 2005). The key data collection requirement for the STS technique is that respondents should provide not a single answer for each item, but a probability range as the basis for the simulation. Specifically, each item in our survey instrument asked for three answers: “minimum likely,” “most likely estimate,” and “maximum likely.” The scaling for these three answers can be of any metric-based type, as long

Table 1. Global Customer-Supplier Relationships Researched.

<table>
<thead>
<tr>
<th>Global Suppliers Researched</th>
<th>Their Global Customers Researched</th>
<th>Global Customers Researched</th>
<th>Their Global Suppliers Researched</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xerox (USA)</td>
<td>ABB (Switzerland)</td>
<td>HSBC (UK)</td>
<td>IBM (USA)</td>
</tr>
<tr>
<td></td>
<td>Siemens (Germany)</td>
<td></td>
<td>NCR (USA)</td>
</tr>
<tr>
<td></td>
<td>Volkswagen (Germany)</td>
<td></td>
<td>British Airways (UK)</td>
</tr>
<tr>
<td></td>
<td>HSBC(^a) (UK)</td>
<td></td>
<td>Lowe(^b) (USA)</td>
</tr>
<tr>
<td>DMV International(^c)</td>
<td>3 major European customers</td>
<td>BG Group(^d) (UK)</td>
<td>Schlumberger (France)</td>
</tr>
<tr>
<td>(The Netherlands)</td>
<td></td>
<td></td>
<td>Bechtel (USA)</td>
</tr>
<tr>
<td>Unilever (UK/The Netherlands)</td>
<td>Wal-Mart (USA)</td>
<td>Siemens (Germany)</td>
<td>2 major U.S. suppliers</td>
</tr>
<tr>
<td></td>
<td>Carrefour (France)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tesco(^e) (UK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WPP(^f) (UK)</td>
<td>Vodafone(^g) (UK)</td>
<td>WPP (UK)</td>
<td>Vodafone (UK)</td>
</tr>
<tr>
<td></td>
<td>Kellogg (USA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BP (UK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Royal Dutch/Shell (The Netherlands/UK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Royal Dutch/Shell</td>
<td>Bosch (Germany)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(The Netherlands/UK)</td>
<td>Daimler-Chrysler (Germany)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unilever (UK/The Netherlands)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wartsila(^h) (Finland)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)The world’s largest bank
\(^b\)One of the world’s largest advertising agencies, and part of Interpublic
\(^c\)Food ingredients, part of Dutch company, Campina
\(^d\)Global energy, particularly gas
\(^e\)largest U.K. food retailer and 4\(^{th}\) largest in the world
\(^f\)The world’s 2nd largest marketing communications company
\(^g\)The world’s wireless communications provider
\(^h\)Power generation and marine propulsion
as the differences between answers have the same meaning (i.e., difference between 1 and 2 is the same as that between 9 and 10).

For the simulation and the subsequent quantitative analysis, we used data from only complete sets of dyads. For the GCM relationships we had two complete sets of surveys for the dyads of DMV (supplier) and a European company (customer) (a GCM relationship which started in 2000), and Xerox (supplier) and Siemens (customer) (a GCM relationship which started in 1990). Each relationship, involving the three different levels of respondents (C1, C2, and C3), were used to create a simulated sample of $n=1,000$, for a total sample size of $n=2,000$ for the GCM analysis.

For the GSM relationships, we obtained data from four complete sets of surveys, for the dyads between BG (customer) and Schlumberger (supplier) (GSM relationship starting date: 2002), HSBC (customer) and IBM (supplier) (GSM relationship starting date: 1997), and HSBC and NCR (GSM relationship starting date: 1990), and BG and Bechtel (GSM relationship starting in late 1990s). Each relationship, involving the three different levels of respondents (S1, S2, and S3), was used to create a sample of $n=1,000$, for a total sample size of $n=4,000$ for the GSM analysis.

### Measures

The appendix contains the measures used in the study. Three participants in each global-account case study (C1, C2, C3, S1, S2, and S3 – described earlier and in Fig. 1) were utilized to obtain the starting values and ranges of those values for each item. Given the phenomena studied and the definitions used for each construct, all of the developed indicators are formative (e.g., Bollen & Lennox, 1991; Jarvis, MacKenzie, & Podsakoff, 2003; MacKenzie, 2003).

In developing formative measures, we followed the guidelines by Jarvis et al. (2003) and treated each latent variable as a “composite factor” where: (1) direction of causality is from the measures to the construct, (2) correlation between the measures assigned to a particular construct may be minimal or even non-existent (i.e., internal consistency is not implied), and (3) dropping an indicator will alter the meaning of the construct. Contrary to reflective measures, where measurement error is accounted for at the indicator level, the measurement error is accounted for at the construct level for formative measures (cf. Bollen & Lennox, 1991). As an added benefit,
the inclusion of formative indicators lends itself well to the STS used to create the large-scale datasets for the GCM \((n = 2,000)\) and GSM \((n = 4,000)\) samples.

The measures of industry globalization drivers (i.e., market-, cost-, government-, and competitive drivers), company global strategy (i.e., market participation, products and services, location of activity, marketing, and competitive moves), and company global organization (i.e., organization structure, management processes, people, and culture) were based on work by Yip (1992) and Johansson and Yip (1994). The measures of demands (i.e., GAM overall, single point of contact, coordination and integration, standardization, consistency, uniform prices, uniform terms of trade, serving in different markets) were adapted from Montgomery and Yip (2000), as were the measures of the extent of a global customer/supplier management program, and the measures of all customers/suppliers. The measure of relative power was developed for this study based on the standard concept in Porter (1980, p. 113). Finally, measures of the three performance items were developed as follows. The measure of sharing of information was based on studies by Lee, So, and Tang (2000), Angulo, Nachtmann, and Waller (2004), and Huang and Gangopadyay (2004). The measures of customer or supplier commitment were based on studies by Provan and Gassenheimer (1994) and Kim (2001). Satisfaction was measured from the perspective of the customer, as recommended by Fornell, Johnson, Anderson, Cha, and Bryant (1996) in their design of the American Customer Satisfaction Index.

**The STS Technique**

Similar to previous studies on supply chains (e.g., Mollenkopf et al., 2005), we used the static simulation approach to generate multiple independent observations for each of the indicators in the appendix based on the sample data provided by the six levels of respondents (i.e., C1, C2, C3, S1, S2, and S3 levels). Specifically, the “triangular generating function” in Microsoft Excel (an add-on statistical technique to the typical Excel software) was used to implement the STS technique. This technique is a part of the “Simtools” packet in Excel which was developed at the Kellogg School of Management, Northwestern University, by Roger B. Myerson. Simtools adds statistical functions and procedures for doing Monte Carlo simulations and risk analysis in Excel.

The variables and associated ranges for each of the six respondent-levels are summarized in Table 2 based on input from the company
Table 2. Variables and Associated Ranges: Average Values of Summated Constructs\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum Likely</th>
<th>Most Likely Estimate</th>
<th>Maximum Likely</th>
</tr>
</thead>
<tbody>
<tr>
<td>IND-C1</td>
<td>5.38</td>
<td>6.75</td>
<td>8.25</td>
</tr>
<tr>
<td>IND-C3</td>
<td>4.38</td>
<td>6.50</td>
<td>7.88</td>
</tr>
<tr>
<td>IND-S1</td>
<td>6.25</td>
<td>7.25</td>
<td>8.25</td>
</tr>
<tr>
<td>IND-S3</td>
<td>3.63</td>
<td>5.25</td>
<td>6.50</td>
</tr>
<tr>
<td>POW-C2</td>
<td>67.50</td>
<td>82.50</td>
<td>97.50</td>
</tr>
<tr>
<td>POW-C3</td>
<td>29.50</td>
<td>34.50</td>
<td>39.50</td>
</tr>
<tr>
<td>POW-S2</td>
<td>36.25</td>
<td>45.00</td>
<td>53.75</td>
</tr>
<tr>
<td>POW-S3</td>
<td>36.77</td>
<td>37.53</td>
<td>40.23</td>
</tr>
<tr>
<td>CUS-C1</td>
<td>15.00</td>
<td>22.50</td>
<td>30.00</td>
</tr>
<tr>
<td>SUP-S1</td>
<td>22.50</td>
<td>30.00</td>
<td>35.00</td>
</tr>
<tr>
<td>STRAT-C1</td>
<td>4.70</td>
<td>6.20</td>
<td>7.80</td>
</tr>
<tr>
<td>STRAT-C3</td>
<td>4.30</td>
<td>6.60</td>
<td>8.10</td>
</tr>
<tr>
<td>STRAT-S1</td>
<td>6.00</td>
<td>7.00</td>
<td>8.00</td>
</tr>
<tr>
<td>STRAT-S3</td>
<td>4.75</td>
<td>6.35</td>
<td>7.65</td>
</tr>
<tr>
<td>ORG-C1</td>
<td>6.50</td>
<td>6.38</td>
<td>8.00</td>
</tr>
<tr>
<td>ORG-C3</td>
<td>4.75</td>
<td>5.50</td>
<td>7.38</td>
</tr>
<tr>
<td>ORG-S1</td>
<td>5.50</td>
<td>6.50</td>
<td>7.50</td>
</tr>
<tr>
<td>ORG-S3</td>
<td>5.81</td>
<td>7.13</td>
<td>8.31</td>
</tr>
<tr>
<td>DEM-C2</td>
<td>4.13</td>
<td>5.50</td>
<td>7.38</td>
</tr>
<tr>
<td>DEM-C3</td>
<td>4.31</td>
<td>6.19</td>
<td>7.44</td>
</tr>
<tr>
<td>DEM-S2</td>
<td>5.47</td>
<td>6.70</td>
<td>8.00</td>
</tr>
<tr>
<td>DEM-S3</td>
<td>4.37</td>
<td>5.67</td>
<td>6.83</td>
</tr>
<tr>
<td>PROG-C1</td>
<td>6.11</td>
<td>7.22</td>
<td>8.33</td>
</tr>
<tr>
<td>PROG-C2</td>
<td>6.22</td>
<td>7.67</td>
<td>8.89</td>
</tr>
<tr>
<td>PROG-C3</td>
<td>2.83</td>
<td>4.39</td>
<td>6.11</td>
</tr>
<tr>
<td>PROG-S1</td>
<td>5.11</td>
<td>7.22</td>
<td>7.67</td>
</tr>
<tr>
<td>PROG-S2</td>
<td>3.89</td>
<td>4.78</td>
<td>5.67</td>
</tr>
<tr>
<td>PROG-S3</td>
<td>5.56</td>
<td>6.75</td>
<td>7.81</td>
</tr>
<tr>
<td>SINFO-C3</td>
<td>2.50</td>
<td>5.50</td>
<td>8.50</td>
</tr>
<tr>
<td>SINFO-S3</td>
<td>5.50</td>
<td>7.50</td>
<td>9.50</td>
</tr>
<tr>
<td>COM-C3</td>
<td>5.00</td>
<td>9.00</td>
<td>12.50</td>
</tr>
<tr>
<td>COM-S3</td>
<td>13.33</td>
<td>25.00</td>
<td>38.33</td>
</tr>
<tr>
<td>SAT-C3</td>
<td>13.33</td>
<td>26.67</td>
<td>31.67</td>
</tr>
<tr>
<td>SAT-S3</td>
<td>20.00</td>
<td>33.33</td>
<td>45.00</td>
</tr>
</tbody>
</table>

\textsuperscript{a}This table shows the average starting values for the summated constructs for the 6 different samples (i.e., C1, C2, C3, S1, S2, and S3) and appropriate scales. The input data for the simulation was taken at the item level, not the construct level. The POW, CUS, SUP, SINFO, COM, and SAT variables were scored as a percentage (%), while all other variables were scored from zero (0) to ten (10).
representatives at the various levels and sides. The data used for input into the simulation consists of a minimum, median, and maximum value on each variable.

Pertinent to our study, the STS simulation modeling approach provides a systems perspective (Scott, 1998) to understanding the interrelationships of the constructs in the global relationship model (cf. McClelland, 1992). Such an approach is particularly useful when trying to assess the relationship effects across a number of variables and multiple levels (cf. Copacino & Rosenfield, 1987). Specifically, since the factor values ranged from low to medium to high, a triangular generating function was used to generate individual observations for the indicators. This technique assumes a cumulative distribution for a triangular probability density function and takes calculated random values for each indicator within the range provided by the respondents at the C1, C2, C3, S1, S2, and S3 levels. In order to ensure a robust statistical power (e.g., Cohen, Cohen, West, & Aiken, 2003; Mollenkopf et al., 2005), 1,000 observations were generated using this function for each of the 6 dyadic global account relationships (i.e., for each of the links of C1-C2-C3 and S1-S2-S3). At a broad level, any number of observations can be generated to ensure adequate power for the statistical analysis to be conducted but we encourage researchers to stay within reasonably “normal” sample sizes for similar research (in most cases the sample size is likely to be between 100 and 2000). Each observation in our sample represents a unique combination of the sets of indicators with each factor independently following a triangular probability distribution.

Fit as Matching

The model includes a series of “fit” relationships between the supplier and customer organizations (i.e., company global strategy, company global organization, demands, extent of a GAM program) as well as between levels (C1 vs. C2 and S1 vs. S2) in an organization (i.e., customization). Based on work by Venkatraman (1989, p. 425), six basic techniques exist to assess fit depending on the (1) “degree of specificity of the functional form of the fit-based relationship” and (2) “number of variables in the fit equation.” Given our focus on criterion-free fits between a relatively small set of variables, the approach best suited to our analyses is “fit as matching.” This perspective is appropriate given the theoretical definition of fit between a limited set of pairs of constructs across levels of informants. The fit between the variables itself is specified without reference to a criterion variable, although subsequently its effect on the three performance variables (SINFO, COM,
and SAT) is assessed. Analytically, we assessed the fit for each of the five variables and levels using an absolute deviation score (e.g., Alexander & Randolph, 1985; Bourgeois, 1985).

**Moderator Hierarchical Regression Analyses**

To test our model, we estimated a multiple hierarchical regression (least squares) model for each of the “customer” (with the C1, C2, and C3 levels) and “supplier” (with the S1, S2, and S3 levels) GRM samples (see Fig. 2). The main effects and moderators were entered in step 1 and the “fit” predictors were entered in step 2.

To avoid potential common methods bias, all measures of explanatory variables were those provided by the GCM or GSM manager side, and all measures of the dependent variables were those taken from the managed side (the customer in the case of GCM and the supplier in the case of GSM). As argued earlier, we believe it is more important to evaluate performance from the viewpoint of those being managed. In addition, as indicated in Fig. 2, some variables used measures from company level respondents, some from account level, and some from an average of the two levels. The fit variables were calculated as differences between sides or levels. Table 3 provides the formulae for the two models.

**Results**

Tables 4 and 5 provide the correlations for the GCM and GSM samples, respectively. Table 6 reports the results of the moderator hierarchical regression testing, including models for each of the three marketing outcome variables (SINFO, COM, and SAT) at the GCM and GSM levels. Since interaction effects were included in the models, the variables were standardized (mean-centered) to avoid potential multi-collinearity problems. The unit of analysis was the global customer account relationship (involving the C1, C2, and C3 levels) for the GCM sample and the global supplier account relationship (involving the S1, S2, and S3) for the GSM sample, respectively.

Each of the six regression models included eight main effects, four moderators, and five “fit” predictors on each of the three marketing outcome variables. The last-step results (step 2) are used to test the hypotheses, while the additional variance explained after step 2 vis-à-vis the variables in step 1 is used to assess the explanatory power of the “fit” variables that were entered into the equation after the main and moderator effects.
Main and Moderator Effects (Step 1)
- Industry – Company Level (C1 or S1)
- Industry – Account Level (C3 or S3)
- Demands – Program Account Level (C2 or S2)
- Demands – Customer/Supplier Account Level (C2 or S2)
- Extent – Average of Corporate & Program Levels (C1+C2 or S1+S2)
- Extent – Customer/Supplier Account Level (C3 or S3)
- Demand (C2 or S2) * Industry (C1 or S1) (Moderator)
- Demand (C3 or S3) * Industry (C3 or S3) (Moderator)
- Extent (C1+C2 or S1+S2) * Industry (C1 or S1) (Moderator)
- Extent (C3+S3) * Industry (C3 or S3) (Moderator)
- Global Customers (C1) or Global Suppliers (S1)
- Relative Power – Account Level (C3/C2 or S3/S2)

Effects of Fit Variables (Step 2)
- Strategy Fit |C1-C3| or |S1-S3|
- Organization Fit |C1-C3| or |S1-S3|
- Customer Demand or Supplier Demand Fit |C2-C3| or |S2-S3|
- Extent of GCM or GSM Program Fit |C12-C3| or |S12-S3|
- Customization of GCM or GSM Program Fit |C1-C2| or |S1-S2|

Marketing Outcomes
- Sharing of Information by Supplier or Customer (C3 or S3)
- Supplier or Customer Commitment (C3 or S3)
- Overall Satisfaction with Supplier or Customer (C3 or S3)

Labels
C1 = Global Customer Account Management – Company Level
C2 = Global Customer Account Management – Account Level Supplier
C3 = Global Customer Account Management – Account Level Customer
S1 = Global Supplier Account Management – Company Level
S2 = Global Supplier Account Management – Account Level Customer
S3 = Global Supplier Account Management – Account Level Supplier

*Fig. 2. Hierarchical Regression Model.*
Table 3. Formulae for Models.

Model for the customer-based (C1-C2-C3) GAM relationship:

\[
Y_{1-3} = \alpha + \beta_1(\text{INDC}_1) + \beta_2(\text{INDC}_3) + \beta_3(\text{DEMC}_2) + \beta_4(\text{DEMC}_3) \\
+ \beta_5(\text{PROGC}_1 + \text{C}_2) + \beta_6(\text{PROGC}_3) + \beta_7(\text{DEMC}_2 * \text{INDC}_1) \\
+ \beta_8(\text{DEMC}_3 * \text{INDC}_3) + \beta_9(\text{PROGC}_1 + \text{C}_2 * \text{INDC}_1) + \beta_{10}(\text{PROGC}_3 * \text{INDC}_3) \\
+ \beta_{11}(\text{CUSC}_1) + \beta_{12}(\text{POWC}_3 / \text{C}_2) + \beta_{13}(|\text{STRATFIT}|) + \beta_{14}(|\text{ORGFIT}|) \\
+ \beta_{15}(|\text{DEMFIT}|) + \beta_{16}(|\text{PROGFIT}|) + \beta_{17}(|\text{ACCTFIT}|) + \epsilon,
\]

\(Y_{1-3}\) = sharing of information by the customer, customer commitment, and satisfaction with the customer
\(\text{INDC}_1\) = industry globalization drivers from the supplier-level corporate perspective
\(\text{INDC}_3\) = industry globalization drivers from the customer-level account’s perspective
\(\text{DEMC}_2\) = demand for a GCM program from the supplier-level account’s perspective
\(\text{DEMC}_3\) = demand for a GCM program from the customer-level account’s representative
\(\text{PROGC}_1 + \text{C}_2, \text{S}_1 + \text{S}_2\) = the extent of a GCM program, as viewed by the average of the company-level (C1) and supplier-level account representatives (C2)
\(\text{PROGC}_3, \text{S}_3\) = the extent of a GCM program from the customer-level account’s representative perspective
\(\text{POWC}_2\) = relative power supplier-level account’s perspective
\(\text{POWC}_3\) = relative power from the customer-level account’s perspective
\(\text{CUSC}_1\) = makeup of all customers from the supplier-level corporate perspective
Table 3. (Continued)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRATFIT</td>
<td>the fit between the supplier’s and customer’s strategies, as viewed by the company-level (C1) and customer-level account representatives (C3)</td>
</tr>
<tr>
<td>ORGFIT</td>
<td>the fit between the supplier’s and customer’s organizations, as viewed by the company-level (C1) and customer-level (C3) account representatives</td>
</tr>
<tr>
<td>DEMFIT</td>
<td>the fit between the supplier’s and customer’s demands a GCM program, as viewed by the supplier-level (C2) and customer-level (C3) account representatives</td>
</tr>
<tr>
<td>PROGFIT</td>
<td>the fit between the supplier’s and customer’s extent of a GCM program, as viewed by the average of the company-level (C1) and supplier-level account representatives (C2) versus the customer-level account representative (C3)</td>
</tr>
<tr>
<td>ACCTFIT</td>
<td>the extent of customization of the GCM program, as indicated by the fit between the views of the company-level (C1) and supplier-level account representatives (C2).</td>
</tr>
</tbody>
</table>

Model for the supplier-based (S1-S2-S3) GAM relationship:

\[
Y_{1-3} = \alpha + \beta_1(INDS_1) + \beta_2(INDS_3) + \beta_3(DEMS_2) + \beta_4(DEMS_3) + \beta_5(PROGS_1+S2) + \beta_6(PROGS_3) + \beta_7(DEMS_2*INDS_1) + \beta_8(DEMS_3*INDS_3) + \beta_9(PROGS_1+S2*INDS_1) + \beta_{10}(PROGS_3*INDS_3) + \beta_{11}(SUPS_1) + \beta_{12}(POWS_3=S2) + \beta_{13}(|STRATFIT|) + \beta_{14}(|ORGFIT|) + \beta_{15}(|DEMFIT|) + \beta_{16}(|PROGFIT|) + \beta_{17}(|ACCTFIT|) + \epsilon,
\]

(Meaning of variables: as for customer-based model above, but read “supplier” for “customer” and vice versa.)
### Table 4. Correlations for the GCM Sample Based on Simulated Data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>IND-C1</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND-C3</td>
<td>.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POW-C2</td>
<td>-.06*</td>
<td>-.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POW-C3</td>
<td>-.86**</td>
<td>-.00</td>
<td>-.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUS-C1</td>
<td>.85**</td>
<td>-.00</td>
<td>-.03</td>
<td>-.98**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRAT-C1</td>
<td>-.85**</td>
<td>-.00</td>
<td>-.02</td>
<td>-.97**</td>
<td>-.96**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRAT-C3</td>
<td>-.83**</td>
<td>-.01</td>
<td>-.03</td>
<td>-.95**</td>
<td>-.94**</td>
<td>.93**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORG-C1</td>
<td>-.57**</td>
<td>-.01</td>
<td>-.02</td>
<td>.66**</td>
<td>-.65**</td>
<td>.63**</td>
<td>.63**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORG-C3</td>
<td>-.83**</td>
<td>-.01</td>
<td>-.01</td>
<td>.95**</td>
<td>-.95**</td>
<td>.93**</td>
<td>.92**</td>
<td>.63**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM-C2</td>
<td>.01</td>
<td>-.04</td>
<td>-.02</td>
<td>.02</td>
<td>.03</td>
<td>-.04</td>
<td>.04</td>
<td>.01</td>
<td>.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM-C3</td>
<td>-.84**</td>
<td>-.01</td>
<td>-.01</td>
<td>.96**</td>
<td>-.96**</td>
<td>.95**</td>
<td>.93**</td>
<td>.64**</td>
<td>.93**</td>
<td>-.03</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROG-C1</td>
<td>-.05</td>
<td>-.04</td>
<td>-.04</td>
<td>-.04</td>
<td>.02</td>
<td>-.00</td>
<td>-.02</td>
<td>.03</td>
<td>-.00</td>
<td>.03</td>
<td>.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROG-C2</td>
<td>-.05</td>
<td>-.04</td>
<td>-.04</td>
<td>-.04</td>
<td>.02</td>
<td>-.00</td>
<td>-.02</td>
<td>.03</td>
<td>-.00</td>
<td>.03</td>
<td>.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROG-C3</td>
<td>-.86**</td>
<td>-.00</td>
<td>.05</td>
<td>.98**</td>
<td>-.97**</td>
<td>.96**</td>
<td>.95**</td>
<td>.65**</td>
<td>.95**</td>
<td>-.07**</td>
<td>.96**</td>
<td>-.01</td>
<td>-.04</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINFO-C3</td>
<td>-.84**</td>
<td>-.01</td>
<td>-.02</td>
<td>.96**</td>
<td>-.96**</td>
<td>.95**</td>
<td>.93**</td>
<td>.64**</td>
<td>.93**</td>
<td>.03</td>
<td>.94**</td>
<td>-.07**</td>
<td>.01</td>
<td>.96**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>COM-C3</td>
<td>-.85**</td>
<td>-.00</td>
<td>-.05</td>
<td>.98**</td>
<td>-.97**</td>
<td>.96**</td>
<td>.94**</td>
<td>.65**</td>
<td>.95**</td>
<td>.05</td>
<td>.96**</td>
<td>-.05</td>
<td>-.03</td>
<td>.98**</td>
<td>.96**</td>
<td>1.00</td>
</tr>
<tr>
<td>SAT-C3</td>
<td>-.84**</td>
<td>-.00</td>
<td>-.03</td>
<td>.96**</td>
<td>-.95**</td>
<td>.94**</td>
<td>.93**</td>
<td>.64**</td>
<td>.93**</td>
<td>-.01</td>
<td>.94**</td>
<td>-.01</td>
<td>-.02</td>
<td>.96**</td>
<td>.94**</td>
<td>.95**</td>
</tr>
</tbody>
</table>

*p < .05.
**p < .01.
Table 5. Correlations for the GSM Sample Based on Simulated Data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>IND-S1</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND-S3</td>
<td>.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POW-S2</td>
<td>-.02</td>
<td>-.81**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POW-S3</td>
<td>-.01</td>
<td>.73**</td>
<td>-.92**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUP-S1</td>
<td>-.02</td>
<td>.01</td>
<td>-.01</td>
<td>.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRAT-S1</td>
<td>.01</td>
<td>.00</td>
<td>.03</td>
<td>.04</td>
<td>-.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRAT-S3</td>
<td>-.02</td>
<td>.27**</td>
<td>.06**</td>
<td>-.24**</td>
<td>-.02</td>
<td>.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORG-S1</td>
<td>-.04</td>
<td>-.01</td>
<td>-.01</td>
<td>.00</td>
<td>.01</td>
<td>.02</td>
<td>-.07*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORG-S3</td>
<td>-.02</td>
<td>-.59**</td>
<td>.80**</td>
<td>-.84**</td>
<td>.02</td>
<td>.01</td>
<td>.29**</td>
<td>-.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM-S2</td>
<td>-.03</td>
<td>-.14**</td>
<td>.85**</td>
<td>-.19**</td>
<td>-.01</td>
<td>.03</td>
<td>.19**</td>
<td>-.00</td>
<td>.17**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM-S3</td>
<td>.02</td>
<td>-.03</td>
<td>.64**</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td>-.03</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.62**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROG-S1</td>
<td>-.04</td>
<td>-.01</td>
<td>-.02</td>
<td>.05</td>
<td>.00</td>
<td>-.03</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td>-.03</td>
<td>-.04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROG-S2</td>
<td>-.01</td>
<td>.05</td>
<td>.93**</td>
<td>.03</td>
<td>-.00</td>
<td>.03</td>
<td>-.02</td>
<td>-.00</td>
<td>-.02</td>
<td>.90**</td>
<td>.62**</td>
<td>-.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROG-S3</td>
<td>-.01</td>
<td>-.50**</td>
<td>.70**</td>
<td>.04**</td>
<td>-.00</td>
<td>.03</td>
<td>-.83**</td>
<td>-.00</td>
<td>-.12**</td>
<td>.81**</td>
<td>.65**</td>
<td>-.02</td>
<td>.95**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINFO-S3</td>
<td>-.05</td>
<td>-.87**</td>
<td>.20**</td>
<td>-.91**</td>
<td>.01</td>
<td>.02</td>
<td>-.22**</td>
<td>.05</td>
<td>.83**</td>
<td>.69**</td>
<td>-.08**</td>
<td>.04</td>
<td>-.01</td>
<td>-.08**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COM-S3</td>
<td>-.03</td>
<td>-.89**</td>
<td>.27**</td>
<td>-.93**</td>
<td>-.00</td>
<td>.00</td>
<td>-.22**</td>
<td>-.01</td>
<td>.85**</td>
<td>.75**</td>
<td>.03</td>
<td>-.00</td>
<td>.01</td>
<td>-.01</td>
<td>.90**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>SAT-S3</td>
<td>.06</td>
<td>-.89**</td>
<td>.90**</td>
<td>-.93**</td>
<td>-.02</td>
<td>-.01</td>
<td>-.22**</td>
<td>.03</td>
<td>.84**</td>
<td>.88**</td>
<td>.00</td>
<td>-.00</td>
<td>-.01</td>
<td>.77**</td>
<td>.43**</td>
<td>.49**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01.
<table>
<thead>
<tr>
<th></th>
<th>Global Customer Management Models ((n = 2,000))</th>
<th>Global Supplier Management Models ((n = 4,000))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SINFO</td>
<td>COM</td>
</tr>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDC₁, S₁</td>
<td>-.03***</td>
<td>-.03***</td>
</tr>
<tr>
<td>INDC₃, S₃</td>
<td>-.01</td>
<td>.00</td>
</tr>
<tr>
<td>DEMC₂, S₂</td>
<td>.00</td>
<td>.03**</td>
</tr>
<tr>
<td>DEMC₃, S₃</td>
<td>.15***</td>
<td>.13***</td>
</tr>
<tr>
<td>PROGC₁+C₂, S₁+S₂</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>PROGC₃, S₃</td>
<td>.51***</td>
<td>.51***</td>
</tr>
<tr>
<td>DEMS₂, C₂ * INDC₁, S₁</td>
<td>.00</td>
<td>-.02*</td>
</tr>
<tr>
<td>DEMS₃, C₃ * INDC₃, S₃</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>PROGC₁+C₂, S₁+S₂* INDC₁, S₁</td>
<td>-.00</td>
<td>-.02</td>
</tr>
<tr>
<td>PROGC₃, S₃* INDC₃, S₃</td>
<td>-.02</td>
<td>-.01</td>
</tr>
<tr>
<td>CUSC₁ or SUPS₁</td>
<td>-.29***</td>
<td>-.32***</td>
</tr>
<tr>
<td>POWC₃/C₂, S₃/S₃</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>.933</td>
<td>.960</td>
</tr>
<tr>
<td><strong>F-value</strong></td>
<td>3847.84***</td>
<td>3840.40***</td>
</tr>
</tbody>
</table>

**Step 2**

|                             |          |          |          |          |          |          |
|-----------------------------|-----------------------------------------------|--------------------------------------------------|
|                             | SINFO    | COM      | SAT      | SINFO    | COM      | SAT      |
| INDC₁, S₁                   | -.01     | -.02**   | -.03***  | -.00     | -.00     | .00      |
| INDC₃, S₃                   | -.01     | .00      | .00      | -.83***  | -.87***  | -.33***  |
| DEMC₂, S₂                   | -.04**   | -.01     | -.05***  | -.14***  | -.07***  | .81***   |
| DEMC₃, S₃                   | .17***   | .14***   | .20***   | -.02     | -.04**   | -.34***  |
Table 6. (Continued)

<table>
<thead>
<tr>
<th>Global Customer Management Models (n = 2,000)</th>
<th>Global Supplier Management Models (n = 4,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SINFO</td>
</tr>
<tr>
<td>PROG(_{C1+C2, S1+S2})</td>
<td>-.04**</td>
</tr>
<tr>
<td>PROG(_{C3, S3})</td>
<td>.60***</td>
</tr>
<tr>
<td>DEM(<em>{S2, C2} \times \text{IND}</em>{C1, S1})</td>
<td>.01</td>
</tr>
<tr>
<td>DEM(<em>{S3, C3} \times \text{IND}</em>{C3, S3})</td>
<td>.02</td>
</tr>
<tr>
<td>PROG(<em>{C1+C2, S1+S2} \times \text{IND}</em>{C1, S1})</td>
<td>-.00</td>
</tr>
<tr>
<td>PROG(<em>{C3, S3} \times \text{IND}</em>{C3, S3})</td>
<td>-.02</td>
</tr>
<tr>
<td>CUS(<em>{C1} \text{or SUP}</em>{S1})</td>
<td>-.15***</td>
</tr>
<tr>
<td>POW(_{C3+C2, S3/S2})</td>
<td>.00</td>
</tr>
<tr>
<td>STRATFIT(_{(C1-C3) \text{or} (S1-S3)})</td>
<td>.00</td>
</tr>
<tr>
<td>ORGFIT(_{(C1-C3) \text{or} (S1-S3)})</td>
<td>-.05***</td>
</tr>
<tr>
<td>DEMFIT(_{(C2-C3) \text{or} (S2-S3)})</td>
<td>.04***</td>
</tr>
<tr>
<td>PROGFIT(_{(C12-C3) \text{or} (S12-S3)})</td>
<td>.07***</td>
</tr>
<tr>
<td>ACCTFIT(_{(C1-C2) \text{or} (S1-S2)})</td>
<td>.00</td>
</tr>
<tr>
<td>(R^2)</td>
<td>.938</td>
</tr>
<tr>
<td>(F)-value</td>
<td>1756.24***</td>
</tr>
<tr>
<td>(\Delta R^2) (from Step 1 to 2)</td>
<td>.005***</td>
</tr>
</tbody>
</table>

* \(p < .10\).
** \(p < .05\).
*** \(p < .01\).
GCM Results
The results for the global customer account management sample are consistent across the SINFO, COM, and SAT models. First, the results show a positive effect of $DEM_C3$ ($\beta_4 = .17$, $p < .01$), $PROG_C3$ ($\beta_6 = .60$, $p < .01$), $DEMFIT$ ($\beta_{15} = .04$, $p < .01$), and $PROGFIT$ ($\beta_{16} = .07$, $p < .01$), and negative effects of $DEM_C2$ ($\beta_3 = -.04$, $p < .05$), $PROG_C1+C2$ ($\beta_5 = -.04$, $p < .05$), $CUS_C1$ ($\beta_{11} = -.15$, $p < .01$), and $ORGFIT$ ($\beta_{14} = -.05$, $p < .01$) on SINFO. The overall SINFO equation had an $R^2 = .938$ ($F$-value = 1756.24, $p < .01$). The “fit” variables explained an additional $0.50\%$ of the variance above the main effects and the moderators ($p < .01$). Second, the results show a positive effect of $DEM_C3$ ($\beta_4 = .14$, $p < .01$), $PROG_C3$ ($\beta_6 = .61$, $p < .01$), $DEMFIT$ ($\beta_{15} = .03$, $p < .01$), and $PROGFIT$ ($\beta_{16} = .07$, $p < .01$), and negative effects of $IND_C1$ ($\beta_1 = -.02$, $p < .05$), $PROG_C1+C2$ ($\beta_5 = -.02$, $p < .05$), $CUS_C1$ ($\beta_{11} = -.19$, $p < .01$), and $ORGFIT$ ($\beta_{14} = -.15$, $p < .01$) on COM. The overall COM equation had an $R^2 = .964$ ($F$-value = 2721.85, $p < .01$). The “fit” variables explained an additional $0.40\%$ of the variance above the main effects and the moderators ($p < .01$). Third, the results show a positive effect of $DEM_C3$ ($\beta_4 = .20$, $p < .01$), $PROG_C3$ ($\beta_6 = .49$, $p < .01$), $DEMFIT$ ($\beta_{15} = .05$, $p < .01$), and $PROGFIT$ ($\beta_{16} = .06$, $p < .01$), and negative effects of $IND_C1$ ($\beta_1 = -.03$, $p < .01$), $DEM_C2$ ($\beta_3 = -.05$, $p < .01$), $CUS_C1$ ($\beta_{11} = -.21$, $p < .01$), and $ORGFIT$ ($\beta_{14} = -.05$, $p < .01$) on SAT. The overall SAT equation had an $R^2 = .931$ ($F$-value = 1580.81, $p < .01$). The “fit” variables explained an additional $0.30\%$ of the variance above the main effects and the moderators ($p < .01$).

GSM Results
The results for the global supplier account management sample are somewhat more diverse than for the customer sample, although the results are largely consistent across the SINFO and COM models but with some varied results for the SAT model. First, the results show positive effects of $PROG_S3$ ($\beta_6 = .12$, $p < .01$) and $ACCTFIT$ ($\beta_{17} = .10$, $p < .01$), and negative effects of $IND_S3$ ($\beta_2 = -.85$, $p < .01$), $DEM_S2$ ($\beta_3 = -.14$, $p < .01$), $PROG_S1+S2$ ($\beta_5 = -.09$, $p < .01$), $DEM_S3 \times IND_S3$ ($\beta_8 = -.05$, $p < .01$), $PROG_S3 \times IND_S3$ ($\beta_{10} = -.08$, $p < .01$), $DEMFIT$ ($\beta_{15} = -.03$, $p < .10$), and $PROGFIT$ ($\beta_{16} = -.11$, $p < .01$) on SINFO. The overall COM equation had an $R^2 = .778$ ($F$-value = 819.92, $p < .01$). The “fit” variables explained an additional $0.50\%$ of the variance above the main effects and the moderators in the SINFO model ($p < .01$). Second, the results show positive effects of $PROG_S3$ ($\beta_6 = .12$, $p < .01$) and $ACCTFIT$ ($\beta_{17} = .09$, $p < .01$), and negative effects of $IND_S3$ ($\beta_2 = -.89$, $p < .01$), $DEM_S2$ ($\beta_3 = -.07$, $p < .01$), $DEM_S3$
Third, the results show positive effects of DEM$_{S2}$ ($\beta_3 = .81$, $p < .01$), PROG$_{S3}$ ($\beta_6 = .56$, $p < .01$), and ACCTFIT ($\beta_{17} = .06$, $p < .01$), and negative effects of IND$_{S3}$ ($\beta_2 = -.35$, $p < .01$), DEM$_{S3}$ ($\beta_4 = -.34$, $p < .01$), PROG$_{S1+S2}$ ($\beta_5 = -.06$, $p < .01$), PROG$_{S3} \times$ IND$_{S3}$ ($\beta_{10} = -.04$, $p < .01$), DEMFIT ($\beta_{15} = -.58$, $p < .01$), and PROGFIT ($\beta_{16} = -.05$, $p < .05$) on SAT. The overall SAT equation had an $R^2 = .763$ ($F$-value $= 755.28$, $p < .01$). The “fit” variables explained an additional 6.40% of the variance above the main effects and the moderators in the SAT model ($p < .01$).

**DISCUSSION**

Our ingoing overall proposition was that the same model might explain the outcomes of GCM and GSM programs. This allowed us to focus more attention on the STS technique which was the main focus of this study to advance strategic management research’s portfolio of research methodologies. Based on the results of the STS technique, a number of implications can be derived. We discuss these implications in an effort to illustrate the use of the STS technique. The outcomes examined are: for GCM, sharing of information and commitment *by* the supplier, and overall satisfaction *with* the supplier, all as perceived by the customer; for GSM, sharing of information and commitment *by* the customer, and overall satisfaction *with* the customer, all as perceived by the supplier.

**Industry Globalization Drivers**

For GCM, industry globalization drivers of the supplier’s industry have small negative effects on marketing outcomes (commitment and satisfaction) while the customer’s industry has no such effect. As a moderator on the extent of demand and extent of programs, drivers have no effect. The most likely interpretation of these results, counter to our hypothesis, is that
in highly global supplier industries (e.g., high technology), customers are harder to satisfy. For GSM, industry globalization drivers of the customer’s industry have zero effect, while drivers for the supplier’s industry have highly negative (−.35 to −.87) effects. As a moderator for the supplier’s view of its own demand for GSM, drivers have a slightly negative effect for two measures, sharing of information (−.05) and commitment (−.05). As a moderator for the extent of the customer’s GSM program, drivers have no effect. As a moderator for the extent of the supplier’s GCM program, drivers have slightly negative effects for all three measures (−.04 to −.08). One interpretation is that, although GSM is more necessary in more global industries, expectations may reduce perceptions of the marketing outcomes. As such, suppliers may need to do even more in terms of GRM in global industries. Overall, industry globalization drivers do not have the predicted positive effect on marketing outcomes on either GCM or GSM, but the effects we do see are similar across the two models.

**Demand for GRM**

For GCM, the supplier’s perception of the customer’s demands has a small negative effect on customer sharing of information (−.04) and overall satisfaction with the customer (−.05). But the customer’s perception of its own demands has moderately positive effects on the three measures of outcomes (0.14–0.20). Hence, more demanding customers perceive greater benefits to themselves from the supplier’s GCM program. For GSM, the customer’s view of its own demand has a negative effect on information sharing (−.14) and commitment (−.07) but highly positive for satisfaction (0.81). Hence, demanding customers perceive greater overall satisfaction by themselves with their suppliers, even if the supplier shares less information and commits less. The supplier’s view of the customer’s demands has a slight negative effect on customer commitment (−.04) and a large effect on overall satisfaction with the customer (−.34). This latter result is not surprising, however. Comparing the two models, we find that while the individual results vary there is a common finding: For both GCM and GSM, more demanding customers are more satisfied with their suppliers. Furthermore, it is the customer’s perception of its own demands that matter, not those of the supplier, i.e., suppliers may well be misreading their customers’ demands.
Extent of GRM Program

For GCM, the supplier’s extent of its GCM program has very small negative effects (−.02 to −.04) on outcomes (as perceived by the customer). For GSM there is a similar effect: the greater the extent of a customer’s GSM, the less happy is the supplier (−.06 to −.09), i.e., a supplier does not like being managed. For GCM, a greater extent of a customer’s corresponding GSM program has highly positive effects (.49 to .61) on its perception of supplier marketing outcome. Similarly, for GSM, a greater extent of a supplier’s corresponding GCM program has a very positive effect (.56) on overall satisfaction with the customer, but this is offset by negative effects (−.12) on customer information sharing and commitment. It seems that the implementers of a GCM or GSM program see little benefit, and some negative effects, from having a greater extent of such programs, which directly contradicts the findings of Birkinshaw et al. (2001). On the other hand, for both GCM and GSM, the corresponding program, from the managed party to manage back, has mostly strong positive effects. Hence, this finding supports the reciprocal view of GRM that there are greater benefits when both sides are prepared for global coordination.

Business with Global Customers or Suppliers

For GCM, the more business accounted for by the customer, the worse the outcomes (−.15 to −.21) of the supplier as perceived by the customer. For GSM, there were no significant effects, i.e., the amount of business accounted for by the supplier made no difference to how the supplier evaluated the customer. It is not surprising that there is a greater negative effect for customer importance than for supplier importance.

Relative Power, Global Strategy Fit, and Global Organization Fit

There are no significant effects for either model regarding these two variables. However, the previous findings imply that it is absolute power rather than relative power that matters. Greater absolute power for customers leads to worse results for the supplier. Similarly, since there are not significant effects for either model for global strategy fit, it seems that higher level global strategy fit between customers and suppliers does not
matter. For GCM, global organization fit has a small to moderate negative effect (−.05 to −.15), while for GSM there are no effects. As such, higher level global organization fit between customers and suppliers does not matter based on our empirical study results.

**Demand Fit**

For GCM, demand fit has slight positive effects (.03 to .05), i.e., a match between a customer’s demands and a supplier’s perception of those demands helps the outcomes of the GCM. For GSM, there are slight negative effects for information sharing and commitment, and a strongly negative effect (−.58) for supplier satisfaction with the customer.

**Program Fit**

For GCM, fit between the supplier’s GCM program and the customer’s corresponding GSM program has slight positive effects (.06 to .07) for the customer’s evaluation of the supplier. In contrast, for GSM, fit between the customer’s GSM program and the supplier’s corresponding GCM program has slight negative effects (−.05 to −.11) for the supplier’s evaluation of the customer.

**Account Customization Fit**

For GCM, customization from the company-level program to the individual customer program has no effect. The implication is that suppliers should stick with a consistent program across the company and not go to the trouble of customization. In contrast, for GSM, there are slight positive effects of customization (.06 to .10). One explanation is that a global supplier deals mostly in one industry with all its customers. Hence, a standard program can be effective, even if the customers are in different industries. For example, an office equipment supplier (like Xerox) can sell its products in the same way to different types of customers. In contrast, a global customer buys from suppliers in many industries. Hence some customization of its GSM can be helpful. For example, a bank (like HSBC in our study) acting as a customer would probably not want to buy computers in the same way that it buys advertising services.
The three variables relating to the broader constructs of globalization – industry globalization drivers, global strategy fit, and global organization fit – did not have significant effects on the outcomes of GCM or GSM programs. It seems that these external (industry) and higher level (company global strategy and organization) constructs have little effect at the program level. What matters is customer demand and the GCM or GSM programs themselves. As between GCM and GSM, it seems that it is easier to have satisfied customers than satisfied suppliers. The various differences between the two sets of models seem to boil down to the fact that suppliers are an unhappy lot, and managing them in a GRM makes them even less happy even though it may benefit the customer. In contrast, GCM programs seem to offer more “win-win” potential. Furthermore, they can be kept relatively simple, without much customization for individual customers. For GSM, the overall implication seems to be that global customers have to be careful in how far they seek to manage (and by implication to squeeze) their global suppliers.

CONCLUSION

This study was developed to introduce the STS technique to the strategic management field. At the same time, this study also reinforces the importance of relationship management programs for achieving favorable outcomes. Our research shows the difficulties and complexities of relationship management, especially in a global context. The STS technique described in this paper lends itself well to the testing of complex strategic management phenomena involving multiple sides of a relationship and multiple levels within organizations. Given that such study situations have been common in the strategic management literature for years (e.g., different levels of leadership in organizations, franchisee–franchisor relationships) and are becoming even more prevalent (e.g., strategic supply chain management), the STS technique offers new avenues to tackle both old and new research questions.

ACKNOWLEDGMENT

We thank The Leverhulme Trust and the U.K.’s Economic and Social Research Council and Engineering and Physical Research Sciences Council for funding.
REFERENCES


**APPENDIX 1. MEASURES**

Basic Instructions Provided to the Survey Participants: Each question will have three answer boxes. These need to be completed with three scores for each question: (1) minimum likely (Min.), (2) most likely estimate (Est.), and (3) maximum likely (Max.). For example, if the question asks to what extent are the “market drivers” favorable for globalization in your industry, your answers might be: 3 for minimum likely, 6 for most likely estimate and 8 for maximum likely.

**IND – Industry Globalization Drivers**

1. Market Drivers (e.g., globally common customer needs and tastes, global customers, global channels, globally transferable marketing, existence of lead countries)

2. Cost Drivers (e.g., global scale economies, steep experience curve effect, sourcing efficiencies, favorable logistics, large differences in country costs, high product development costs, fast changing technology)
3. Government Drivers (e.g., favorable trade policies, compatible technical standards, common marketing regulations, government-owned competitors, absence of government owned customers, low importance to national governments)

4. Competitive Drivers (e.g., high exports and imports, competitors from different continents, interdependence of countries, competitors globalized, transferable competitive advantage)

DEM – Demands (0 = “do not demand at all” to 10 = “demand very much”)

1. Global account management overall
2. A single point of contact
3. Greater global coordination and integration of resources for serving customers
4. Greater standardization across countries in your products or services
5. More consistency in service quality and performance
6. More uniform prices charged to them in the different countries in which you serve them
7. More uniform terms of trade (other than price) in the different countries in which you serve them
8. Serving them in a market in which you do not or did not have operations

PROG – Extent of Global Customer Management Program (0 = “not use at all” to 10 = “use very much”)

1. Managers, directors, or similar positions responsible for specific global accounts
2. Support staff or team for the global accounts
3. Customer information system
4. Revenue/profit measures for the GCM program
5. Reporting processes for the GCM program
6. Evaluation of the personnel involved on GCM objectives
7. Global personnel incentives and compensation
8. Global marketing plan
9. Customer councils or panels

PROG – Extent of Global Supplier Management Program (0 = “not use at all” to 10 = “use very much”)

1. Managers, lead buyers, or similar positions responsible for specific global suppliers
2. Support staff or team for the global suppliers
3. Global supplier information system
4. Revenue/profit measures for the GSM program
5. Reporting processes for the GSM program
6. Evaluation of the personnel involved
7. Global personnel incentives and compensation
8. Global procurement plan
9. Supplier councils or panels

CUS – Business with Global Customers (percentage)

1. What percentage would you consider to be multinational customers? i.e., those who buy from you in more than one country regardless of whether they coordinate purchases across countries?
2. Approximately what percentage of your revenues is accounted for by coordinated global customers, i.e., those who buy from you in more than one country who coordinate purchases across countries?

SUP – Business with Global Suppliers (percentage)

1. What percentage would you consider to be multinational suppliers? i.e., those who sell to you in more than one country regardless of whether they coordinate sales across countries?
2. Approximately what percentage of your purchases is accounted for by coordinated global suppliers, i.e., those who sell to you in more than one country who coordinate sales across countries?

POW – Relative Power, Global Customer Management Program (percentage)

1. Of the product category for which you are responsible, what percentage of your total business is accounted for by this customer?

POW – Relative Power, Global Supplier Management Program (percentage)

1. Of the product category for which you are responsible, what percentage of your purchases is accounted for by this supplier?

STRAT – Company Global Strategy (0 = “not use at all” to 10 = “use very much”)

1. Global market participation (e.g., global market share, presence in globally strategic countries)
2. Global products and services (e.g., globally standardized or uniform)
3. Global location of activities (e.g., a globally concentrated and/or coordinated network of activities such as manufacturing or production)
4. Global marketing (e.g., globally uniform elements of marketing mix)
5. Global competitive moves (e.g., multi-country competitive moves, counterparty moves, globally coordinated moves)

ORG – Company Global Organization (0 = “not use at all” to 10 = “use very much”)

1. Global organization structure (e.g., global business or functional heads, global management teams)
2. Global management processes (e.g., strategic information system, cross-country coordination, knowledge sharing, strategic planning, budgeting, customer management, performance review and compensation)
3. Global human resources (e.g., global allocation of personnel, global board of directors)
4. Global culture

MKTOUT – Marketing Outcomes, Global Customer Management Program (percentage change), as perceived by the customer

1. Sharing of information by the supplier, % change since program started (SINFO)
2. Supplier commitment, % change since program started (COM)
3. Your overall satisfaction with the supplier, % change since program started (SAT)

MKTOUT – Marketing Outcomes, Global Supplier Management Program (percentage change), as perceived by the supplier

1. Sharing of information by the customer, % change since program started (SINFO)
2. Customer commitment, % change since program started (COM)
3. Your overall satisfaction with the customer, % change since program started (SAT)
APPENDIX 2. DEVELOPMENT OF THE HYPOTHESES

Two types of relationships are tested via the STS technique in this study: direct effects and fit effects. Each research hypothesis is developed below based on the relevant literature.

Direct Effects

Industry Globalization Drivers
Some researchers stress that industry characteristics are a key determinant of whether companies should use global, as opposed to multi-local, strategy (Porter, 1986; Yip, 1989). Yip and Madsen (1996) and Montgomery and Yip (2000) extend this view of industry globalization drivers to apply to GCM as a type of global strategy. Specifically, industries with strong forces for globalization – such as globally common customer needs, global scale economies, few government barriers to trade and investment, and drivers for global competition (Yip, 1992) – will require companies to respond with globally integrated strategies, including GCM programs. Similarly, customers in industries with strong globalization drivers will also need to respond with global strategies that include GSM programs. Hence, strong industry globalization drivers should have a positive effect on performance improvement from having a GRM program.

H1a. Stronger globalization drivers in a supplier’s industry result in better marketing outcomes from a GCM program.

H1b. Stronger globalization drivers in a customer’s industry result in better marketing outcomes from a GSM program.

Demand for GRM Services
Implementing GRM programs is an onerous activity (e.g., Nahapiet, 1994) that will hurt the provider’s financial and other aspects of performance unless there are positive outcomes. Such outcomes should be driven particularly by whether the recipient of the services demands them and finds them of benefit. This dependence (of the performance consequences of the GRM provision) on demand very much accords with contingency theory (e.g., Lawrence & Lorsch, 1967; Miles, Snow, & Pfeffer, 1974), with standard microeconomic supply and demand theory, and with relationship
marketing theory (e.g., Grönroos, 1997; Kalwani & Narayandas, 1995; Morgan & Hunt, 1994). Even so, in regard to GCM, there is evidence that some suppliers resist it because of their fear that its adoption will help the customer demand lower prices. For example, Yip and Madsen (1996) cite Xerox as denying customers’ requests for GAM if the request is motivated only by a desire to pay uniform prices worldwide.

**H2a. Greater demand by customers for GCM services results in better marketing outcomes from a GCM program.**

In regard to GSM, suppliers themselves would not demand a GSM program, as it is the customer’s prerogative whether to have one. Hence, we focus on the extent to which a customer demands that its suppliers interact with its GSM program in a globally coordinated way by also providing GRM services.

**H2b. Greater demand of customers with a GSM program for GRM services from their suppliers results in better marketing outcomes from a GSM program.**

**Use of GRM Program Aspects**

Although implementing GRM programs can be onerous, both corporate experience and some research show that if done right there can be significant benefits to the adopter. In particular, for GCM programs, Birkinshaw et al. (2001) found that greater benefits were associated not with particular aspects of these programs but with the extent to which the adopter used all aspects. As such, the more aspects of a program that were adopted the greater the benefits turned out to be for the participants. For GSM, the greater ease by which customers can impose a GSM program on their suppliers means that there may be an even stronger effect of using GRM aspects in GSM than in GCM. In the latter case, suppliers depend on the compliance and acceptance of their customers.

**H3a. More extensive use by a supplier of aspects of a GCM program results in better marketing outcomes from a GCM program.**

**H3b. More extensive use by a customer of aspects of a GSM program results in better marketing outcomes from a GSM program.**
Extent of Business with Global Customers or Suppliers
The more business that a company does with global customers or suppliers, the greater its need for a GRM program, and the greater should be the effect on performance improvements from having a GRM program. Operationally, we consider that the single best indicator of the globality of a company’s customers and suppliers is the percentage of a company’s sales to globally coordinated customers, or the percentage of purchases from globally coordinated suppliers.

H4a. More business for a supplier with global customers results in better marketing outcomes from a GCM program.

H4b. More business for a customer with global suppliers results in better marketing outcomes from a GSM program.

Relative Power
Relative power of the parties in a relationship affects performance in buyer–seller dyads. Indeed, bargaining power relative to suppliers and customers constitute two of Porter’s five forces (Porter, 1980). Following Porter, we argue that a supplier has greater power relative to a global customer if the percentage of its sales (in a given category) to the customer is greater than the percentage of the customer’s purchases (in the same category) from that supplier. The parallel argument holds for a customer and a global supplier to that customer. In turn, this greater power should result in greater performance improvement from having a GRM program.

H5a. Greater relative power for a supplier results in better marketing outcomes from a GCM program.

H5b. Greater relative power for a customer results in better marketing outcomes from a GSM program.

Fit Relationships
Company Global Strategy and Global Organization Fit
Research on relationship marketing (e.g., Grönroos, 1997; Kalwani & Narayandas, 1995) shows that greater fit in strategic objectives between suppliers and their customers improves relationships. Birkinshaw et al. (2001) argue that GCM programs work better when both sides have similar, preferably high levels of global integration capability. GRM programs fit in
the broader context of a company’s global strategy and organization or extent of global integration. Hence, we need to look at company (or business unit) level use of global strategy and global organization, as established in the global strategy literature (e.g., Bartlett & Ghoshal, 1989; Yip, 1992; Kim & Mauborgne, 1995; Westney & Zaheer, 2001). In essence, company use of global strategy involves business choices such as globally standardized products, global location of activities, and global marketing. Company use of a global organization framework involves choices such as global organization structures, global management processes, and global human resource policies. GRM programs on either side of the supplier–customer dyad work better internally if within the context of higher levels of global integration in terms of strategy and organization. Hence, similarity in company-level global strategy and organization between supplier and customer will enhance the performance of GRM programs.

**H6. Closer fit between a supplier’s and a customer’s company global strategies results in better marketing outcomes from either’s GRM program.**

**H7. Closer fit between a supplier’s and a customer’s company global organization results in better marketing outcomes from either’s GRM program.**

**Demand Fit**

Contingency theory would argue that a fit between customer demand for GCM services and supplier provision of such services would improve performance for both the supplier and the customer. Montgomery and Yip (2000) offer some evidence that suppliers do provide more GAM services in response to customers’ demand. However, their study did not use the same variables for both demand and supply and they collected data from the viewpoint of only the supplier. At the same time, before even providing a set of GCM services for a particular customer, the supplier needs to perceive the customer’s demands correctly. Hence, we propose that a more complete test would examine the match, on the same variables, for demand by the customer and perception of that demand by the supplier. Montgomery and Yip (2000) identified these elements of GCM demand as single point of contact, coordination of resources for serving customers, uniform prices, uniform terms of trade, standardization of products and services, consistency in service quality and performance, and service in markets in which supplier has no operations.
**H8.** A closer fit between a customer’s demand for GRM services and a supplier’s perception of that demand results in better marketing outcomes from either’s GRM program.

*Programs Match*
Both contingency theory and theories about interorganizational isomorphism (Brouthers et al., 2005; Nickerson & Silverman, 2003) would argue that interorganizational relationships work better if both sides of the dyad structured themselves in the same way to manage each other. In the case of GRM, this would mean parallelism between a customer’s supplier management program (GSM) and a supplier’s customer management program (GCM). Montgomery and Yip (2000) identified these elements of GCM program supply as global account managers, support staff, revenue or profit measures, reporting processes, customer information, personnel evaluation, incentives and compensation, and customer councils or panels.

**H9.** A closer fit between a supplier’s and its customer’s GRM programs results in more positive performance improvements from either’s GRM program.

*Program Customization*
Nearly all companies with GCM or GSM programs apply these to more than just one customer or supplier. Hence, companies face the challenge of finding a balance between standardization across accounts and customization for individual accounts. Standardization brings the (mostly internal) benefits of scale and simplicity, while customization brings the (mostly external) benefits of better fit for the account’s needs. The issue of standardization versus customization across customers should not be confused with the more usual discussion in the global strategy literature of standardization or customization across countries. Indeed, within a single global account, a critical issue is the extent to which there should be cross-country standardization. On balance, we expect that greater customization of a GRM program would result in better performance, especially when perceived at the account level, whether by the customer or the supplier.

**H10.** Greater customization of the company level GRM program, for the individual account level, for either suppliers or customers, results in better marketing outcomes for either’s GRM program.
NEW METHODS FOR EX POST EVALUATION OF REGIONAL GROUPING SCHEMES IN INTERNATIONAL BUSINESS RESEARCH: A SIMULATED ANNEALING APPROACH

Paul M. Vaaler, Ruth V. Aguilera and Ricardo Flores

ABSTRACT

International business research has long acknowledged the importance of regional factors for foreign direct investment (FDI) by multinational corporations (MNCs). However, significant differences when defining these regions obscure the analysis about how and why regions matter. In response, we develop and empirically document support for a framework to evaluate alternative regional grouping schemes. We demonstrate application of this evaluative framework using data on the global location decisions by US-based MNCs from 1980 to 2000 and two alternative regional grouping schemes. We conclude with discussion of implications for future academic research related to understanding the impact of country groupings on MNC FDI decisions.
INTRODUCTION

International business (IB) research on determinants of country attractiveness for foreign direct investment (FDI) by multinational corporations (MNCs) has long emphasized the importance of a country’s regional grouping. Countries in the “Triad” of North America, Western Europe and Greater Japan (Ohmae, 1985; Rugman & Verbeke, 2004) are considered more attractive for investment than similarly situated countries outside this Triad. Countries from Latin America with civil law traditions are less attractive for lending and investment than South Asian countries with common law traditions (La Porta, Lopez-de-Silanes, Shleifer, & Vishny, 1998). Alternative regional groupings based on cultural affinity (Hofstede, 2001 [1980]; Ronen & Shenkar, 1985), geo-political orientation (UN, 2007) and/or economic development levels (Vaaler & McNamara, 2004) are also considered important in understanding individual country attractiveness for MNC FDI.

But this IB research stream faces challenges related to the ex ante grounding of these regional groupings. Sometimes their justification follows from research intuition, bald assumption or anecdotal support. Even when grounded in theory, regional groupings are often vulnerable to reasonable refinements that can substantially change their power to explain important IB phenomena, including but not limited to MNC FDI. These concerns undermine regional grouping concepts, constructs and measures and impair the validity and reliability of theoretical and empirical IB research relying on them. In this context, we see an opportunity to contribute methodologically with an alternative approach to assessing regional grouping schemes. This approach complements ex ante theoretical assessment of regional grouping schemes with ex post assessment of their robustness to reasonable refinement. It promises insight on current and future IB research regarding the absolute and comparative robustness of different regional grouping schemes important to the study of MNC FDI patterns and broader questions of individual country attractiveness for lending and investment.

In the next section, we begin to develop these points, first by surveying the theoretical and practical grounding of alternative regional grouping schemes used in recent IB research. We hold that some grouping schemes lack ex ante theoretical grounding and follow from ad hoc researcher assertion or intuition. Alternatively, ex ante theoretical grounding is often weak, thus rendering schemes vulnerable to substantial change after reasonable refinement. In this context, we propose an ex post empirical technique for complementary evaluation of alternative grouping schemes using a novel algorithm based on simulated annealing.
Following the next section we describe in greater detail that, simulated annealing permits iterative refinement and optimization of initial grouping schemes where the number of alternative grouping schemes is great and global optimization of the scheme based on desired criteria is challenged by the existence of several local optima (Fox, Srinivasan, & Vaaler, 1997; Goffe, Ferrier, & Rogers, 1994). Given an initial regional grouping scheme and some theorized relationship between the grouping scheme and some phenomenon of IB research interest, we can then iteratively estimate, refine and then re-estimate the impact of alternative grouping schemes as the algorithm heads toward the global optimum. Comparison of differences in grouping schemes before and after simulated annealing provides the basis for ex post evaluation of regional grouping scheme robustness. Schemes with more (less) change in the number of groups, more (less) change in the sign and significance of non-group factors, and more (less) change in overall regression equation explanation of variation in phenomenon of interest are less (more) effective at supporting research inferences of interest.

In the penultimate section, we illustrate our approach to ex post evaluation of alternative regional grouping schemes using recent empirical analyses reported in Flores and Aguilera (2007). They estimate the likelihood of US MNC FDI in countries around the world in 1980 and 2000 using regression equation with several country-specific economic and cultural factors as well as regional group dummies. We re-create their analyses using two different regional grouping schemes: (1) regional grouping based on continental location (North American, South American, Europe, Africa, Australia-Asia) and (2) regional grouping based on a scheme proposed by Vaaler and McNamara (2004) to explain differences in country sovereign risk (North America, Latin America-Caribbean, Western Europe, Central-Eastern Europe, Africa-Middle East, Australia-Asia). After initial Logit estimation of the likelihood of US MNC investment in foreign countries using these country specific and one of these two alternative regional grouping schemes, we submit the initial schemes to simulated annealing analysis. At the conclusion of this analysis, we compare the extent of before and after change for each scheme and assess robustness of each scheme alone and in comparison.

Lastly, we summarize the central issue and findings of this methodological research paper. We note several implications for IB and related management research reliant on the validity and reliability of groups, whether they are regionally, or otherwise defined. We propose practical strategies for implementing such ex post schemes for evaluating schemes and suggest how simulated annealing itself might be incorporated into future empirical work.
REGIONAL GROUPING SCHEMES IN IB RESEARCH

How and why are regional grouping schemes important in IB research? To address this question and set the context for our survey of alternative grouping schemes we rely primarily on Aguilera, Flores, and Vaaler (2007), who answer this question in detail in a companion paper. The last two decades of research in IB and related management fields have seen substantial debate about the significance and relative importance of country location for understanding its attractiveness for MNC investment. Democratization in local polities, privatization and deregulation in local economies, as well as international regimes promoting trade liberalization have all promoted the position of MNCs as instruments of country investment and growth, as well as instruments of regional integration and, indeed, globalization of formerly segmented national markets (Dicken, 1998; Giddens, 1999; Held, 2000).

Yet, research on central tendencies in MNC internationalization remains inconclusive and requires more systematic analysis. On the one hand, IB scholars such as Rugman and Verbeke (2004, 2007) hold that MNCs locational patterns have become increasingly regional as opposed to global. Thus, understanding the impact of regional country groupings is increasingly important for explaining whether and how MNCs internationalize. They contrast with other IB scholars who emphasize the value of global scale and scope in MNC operations (Agmon, 2003; Bird & Stevens, 2003; Clark, 1997; Clark & Knowles, 2003; Clark, Knowles, & Hodis, 2004). Regional patterns of operation are still important to study from this alternative view, however, as they represent intermediate steps in the internationalization trajectories of firms transforming themselves from domestic to regional to worldwide competitors. Thus no matter the side researchers take in this debate, understanding alternative regional grouping schemes and their impact on MNC internationalization behavior becomes critical. We survey those schemes and raise issues related to their prudential use, validity and reliability in recent IB research.

Defining Regional Groups

In this context of debate over MNC internationalization patterns and the impact of regions, we see value first in seeking to define the concept of
central interest. The term region might be intuitively defined as a “fairly large area of a country or of the world, usually without exact limits” (Longman, 1995). This definition implies proximity between countries based on physical dimensions of measurement. As we will see below, however, scholars in IB and related fields have defined regions by alternative dimensions of proximity. Countries have been grouped based on broad patterns of trade and economic relationships (e.g., Rugman & Verbeke, 2004), based on broad cultural indices (e.g., Hofstede, 2001 [1980]), based on key components of culture such as language, religion, law, politics and popular media (e.g., Guiso, Sapienza, & Zingales, 2006) as well as on shared physical proximity (e.g., Vaaler & McNamara, 2004). We survey some such schemes and note their key findings related to debate over MNC internationalization patterns.

Regional Grouping Schemes Based on Economics and Trade

Though readers might intuit that physical proximity is the most common dimension for grouping, dimensions related to common levels of economic development tend to dominate most schemes in IB research. Several studies stress the need of looking at the outcomes of regional economic integration (Frankel, 1997). One of the forerunners of this approach was Ohmae (1985), who grouped countries into a Triad of three regions centered on Japan, the US and Western Europe, primarily France, Germany, and the UK. He claimed that MNC survival required some dominant market positioning in at least one these three national economies, and by implication, the North American, European and/or Asian countries that depended on each Triad leader.

Building on Ohmae’s insights, Rugman and Verbeke note that regional FDI by MNCs follow multilateral trade regimes such as the North American Free Trade Agreement (NAFTA), the Association of Southeast Asian Nations (ASEAN) and the European Union (EU) (Rugman, 2005; Rugman & Verbeke, 2004, 2005). Researchers have highlighted the relevance of countries’ membership in key transnational organizations such as the Organisation for Economic Co-operation and Development (OECD) (Buckley & Ghauri, 2004; Dunning, 2001; Gatignon & Kimberly, 2004). On the other hand, an emerging literature in political economy suggests that regional FDI follows a more complex regional grouping based on multilateral regimes and bi-lateral investment treat arrangements (Simmons, Elkins, & Guzman, 2006). Thus, regional trading blocs and economic
arrangements might benefit from refinement based on additional bilateral dyads and arrangements.

Regional Grouping Schemes Based on Culture

The most common regional criteria used by scholars grouping countries accordingly focus on cultural dimensions related to the personal attitudes and beliefs.¹ Perhaps the prominent application of cultural dimensions to group countries together for explanation of MNC behavior comes from Hofstede (2001 [1980]). He first surveyed IBM employees in the 1970s to derive cultural dimensions related to 53 countries. Relying on a statistical technique (hierarchical clustering) for the cultural dimensions he uncovers in his studies (power distance, uncertainty avoidance, masculinity/femininity and individualism/collectivism), Hofstede ended up defining a 12-group regional structure (Hofstede, 2001 [1980], p. 62).

Hofstede’s indices have provided the basis for subsequent empirical studies that have documented similarity (dissimilarity) between MNC investment and competitive behaviors within (between) regions defined by different factors and clusters (see Kirkman, Lowe, & Gibson, 2006 for a complete review of the consequences of Hofstede’s framework). Later, Ronen and Shenkar (1985) offered their own scheme, partially using the work of Hofstede, where 45 countries were grouped into nine cultural clusters, while Furnham, Kirkcaldy, and Lynn (1994) offered their own scheme of 41 five cultural clusters. More recently, the World Values Survey (Abramson & Inglehart, 1995), is finding more use in IB research. The so-called GLOBE project represents yet another stream flowing from Hofstede (House, Javidan, Hanges, & Dorfman, 2002). Gupta, Hanges, and Dorfman (2002) have used GLOBE project data in discriminant analyses to identify 7 regional groups for 61 countries involved in the GLOBE project, while Brodbeck and a large team of researchers in European countries (Brodbeck et al., 2000) and Lenartowicz and Johnson (2003) in Latin America have used GLOBE project data to identify intra-regional grouping schemes relevant to MNC behavior.

Regional Grouping Schemes Based on Institutions

Yet another approach to using cultural dimensions relies less on aggregate indices and more on specific cultural traits such as language, religion, law,
politics and media. This approach comprises both culture defined by individual attitudes and beliefs and culture-as-institutions, that is, the collective legal, political and social arrangements that spring from such attitudes and beliefs and together guide basic rules of economic exchange (North, 1990). Language and religion are particularly important cultural components, such as in work by Chetty, Eriksson, and Lindbergh (2006), Dow and Karunaratna (2006) and Leung, Bhagat, Buchan, Erez, and Gibson (2005).

This approach contrasts with regional groupings based explaining MNC FDI and lending based on similar levels of economic development (Dunning, 1998, 2001), based on similar levels of corruption, bureaucratic efficiency, media and voice, respect for law (Globerman & Shapiro, 2003; Kaufmann, Kraay, & Zoido-Lobaton (1999); La Porta, Lopez-de-Silanes, Shleifer, & Vishny, 1999). La Porta and his colleagues, for example, document that countries with Anglo-American common law traditions providing stronger investor and creditor protections draw more foreign investment, have deeper and broader debt markets compared to countries with French civil law traditions. Aguilera and Cuervo-Cazurra (2004) note that their findings support the idea that countries with more protective (of minority shareholder rights) legal systems tend to develop stronger and better-enforced codes of good corporate governance. On the other hand, Berkowitz, Pistor, and Richard (2003) show that many results obtained by La Porta and his colleagues vary once more refined country groupings are defined. Berkowitz and colleagues distinguish between common and civil law countries where the legal system was imposed by force or developed organically. Countries where legal system developed organically, whether civil or common law in nature, provide more protection than in countries where the system was forcibly “imported.”

Regional Grouping Schemes Based on Geography

Physical proximity and contiguity present the most straightforward dimensions for creating regional grouping schemes. Here, shared geography overlaps with and contributes to other similarities along dimensions previously surveyed above. Dividing the world into continental groupings such as Europe, Asia, America, Africa and Oceania often appears in IB research. Kwok and Tadesse (2006) choose continental groupings to study the free market-orientation of financial systems in 41 countries. Similarly, Katrishen and Scordis (1998) find that the continent from which MNC
insurers are domiciled is significantly linked to MNC insurer likelihood of achieving economies of scale. Geringer, Beamish, and daCosta (1989) also control by continent of origin when assessing performance of 200 MNCs with differing levels of diversification and internationalization. Vaaler and McNamara (2004) find that continental regional specialization by major credit rating agencies significantly and substantially changes their sovereign risk assessments in the late 1980s and 1990s.

The United Nation’s Statistics Division may offer more fine-grained partition of these geographic regions (UN, 2007). The UN scheme breaks up countries into 19 regions (i.e., Australia and New Zealand, Caribbean, Central America, Eastern Africa, Eastern Asia, Eastern Europe, Melanesia, Middle Africa, Northern Africa, Northern America, Northern Europe, South America, South-Central Asia, South-Eastern Asia, Southern Africa, Southern Europe, Western Africa, Western Asia, Western Europe). Flores and Aguilera (2007) use this scheme to explain US MNC country location decisions in 1980 and 2000.

Use, Validity and Reliability Issues in Recent IB Empirical Research

Our review of the empirical literature related to regional effects in IB and related fields reveal clear differences regarding dimensions for grouping countries and explaining MNC behavior and performance within and across such groupings. No doubt this often follows from the eclecticism of IB research interests, theoretical perspectives and empirical analytical methods. Even so, some such dimensions are provided without any ex ante theoretical grounding, thus undermining concept, construct and measurement validity. Even where ex ante theoretical grounding is provided, we note in many cases that alternative schemes based on similar theories and methods yield different results, thus impairing reliability claims as well.

For example, Ronen and Shenkar (1985) refine Hofstede’s regional clusters with differing results regarding MNC executive attitudes, while Simmons et al. (2006) suggest that refinement of regional trading blocs based on assessment of bilateral investment treaties may change previous results based on multilateral trade agreements alone. And more fine-grained regional grouping schemes based on legal system differences reported by Berkowitz et al.(2003) yield different insights on the extent of investor and creditor protection for MNCs compared to more coarse-grained measures and groupings proposed by La Porta and his colleagues.
A NEW EX POST APPROACH TO ASSESSING REGIONAL GROUP VALIDITY AND RELIABILITY: SIMULATED ANNEALING

Ex Post Approaches to Assessing Regional Grouping Schemes

We take such findings as the departure point for our own alternative approach to evaluation of regional grouping schemes. Rather than attack the theoretical validity and reliability of prominent regional grouping schemes reliant on economic, cultural, institutional and other dimensions ex ante, we propose an alternative ex post evaluative technique based on iterative refinement and re-estimation using a simulated annealing approach. However, the regional scheme is structured, the ex post question to be posed is whether and how much such a scheme is subject to change after reasonable refinement. Schemes that are vulnerable to substantial change have weaker validity and reliability than schemes exhibiting less change. Evidence of robustness after submission to this ex post evaluation responds to criticisms noted above with empirical evidence demonstrating the stability of key grouping assumptions grounded in whatever theory IB researchers choose initially to justify their regional grouping scheme.

Consider, for example, an empirical model of MNC country location defined as follows:

\[ \text{MNC subsidiary}_{ijmt} = \alpha_0 + \sum_{k=1}^{l} \text{Country factors}_{i_k} + \sum_{n=1}^{p} \text{MNC factors}_{j_n} + \sum_{q=1}^{r} \text{Year factors}_{t_q} + \sum_{s=1}^{u} \text{Regional dummies}_{m_s} + e_{ijmt} \]  

In (1), the dependent variable is a 0-1 indicator equal to 1 when MNC \( j \) has a subsidiary operation in country \( i \) part of region \( m \) in year \( t \). We explain the likelihood of MNC location of a subsidiary operation based on country factors \( i \) \((k=1-l)\), MNC factors \( j \) \((n=1-p)\) and time (year) factors \( t \) \((q=1-r)\). In addition, we define a regional grouping scheme in the form of fixed regional dummies \( m \) \((s=1-u)\). The structure of this regional grouping scheme is presumably grounded ex ante in theory related to the significance of economic, cultural, institutional and/or geographic factors. Logistic or Probit estimation of this model provides insight on the impact of regional grouping based on evaluation of the regional dummies for their individual and collective significance and practical impact.
Our approach implies re-estimation of (1) after iterative ex post refinement of the initial regional grouping schemes. This implication raises new challenges related to the extent of this refinement. In concept, alternative regional groupings are limited only by the number of countries and potential country combinations. It is unfeasible to search all of these possible alternative schemes. A partial search seeking to refine the grouping scheme based on some simple optimization criterion may reduce search time, yet challenges still persist. Consider, for example, a search to refine some initial regional grouping scheme based on minimization of the regression equation’s unexplained variance, that is, the error sum of squares (ESS) generated by logistic estimation. If the number of alternative grouping schemes with refinement is still large, simple minimization using conventional algorithms such as Newton Raphson or Davidson-Fletcher-Powell is likely to move greedily to a local minimum but search no further. Thus, we may end search and refinement of initial regional grouping scheme prematurely, thus leaving the global minimum ESS unidentified and the researcher unsure as to the stability of initial results.

**Ex Post Evaluation Based on Simulated Annealing**

An alternative “simulated annealing” search algorithm improves on these and other “hill-climbing” heuristics. Usually the cooling process for molten metal is used to detail how this procedure works. On those processes where the temperature of the metal is continuously reduced, after a slow cooling (annealing), the metal arrives to a minimum energy state. Innate random variations in energy allow the annealed system to escape local energy minima. Even though this technique is not flawless in finding the global minimum, it tends to achieve an ending point closer to the global minimum than do conventional algorithms (Alrefaei & Andradóttir, 1999; Goffe et al., 1994). Perhaps the best-known application of simulated annealing is to the “traveling salesman” problem, where the goal is to find the minimum trip distance connecting several cities. Academic applications of this technique range from optimal land use and irrigation design (Aerts & Heuvelink, 2002) to micro-circuit design (Kirkpatrick, Gelatt, & Vecchi, 1983). Within the management realm, Han (1994) uses simulated annealing for optimal information filing, while Carley and Svoboda (1996) model optimal organizational adaptation to environmental shocks. Semmler and Gong (1996) optimize the size of industry groupings in analyses of real business cycle parameters, while Fox et al. (1997) use simulated annealing to refine
business membership in standard industry classes and assess the impact of such intra-industry strategic groups on business performance in the US during the 1970s.

To explain how the annealing algorithm functions in our application, consider an initial partitioning of countries into regional groups based on (1): \( \{ P_s \} = (p_{s=1}, p_{s=2}, \ldots, p_{s=U}) \). Here, \( p_s \) represents the \( s \)th regional group composed of \( n \) countries. Coefficients are estimated for this initial partition. Next, a new partition \( \{ P_s' \} \) is made by varying the group structure of the whole set of countries. The variation may be of two types:

1. It may be a random exchange of two countries from different regional groups, \( p_s \).
2. It may be a random perturbation changing the size of a given region, \( p_s \), resulting in a change in the number of countries \( n \) in the region from \( x \geq 3 \) to \( x - c \geq 3 \) where \( c \) is some integer.

After re-estimation, if the new ESS’ is less than the old ESS, the new regional group structure, \( \{ P_s' \} \), structure replaces the old regional group structure, \( \{ P_s \} \) and the algorithm moves downhill. If the new ESS’ is greater than or equal to the old ESS, then acceptance is stochastic. A criterion developed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) decides on acceptance of an uphill move. Thermodynamics analogies also motivated the Metropolis criterion. The value:

\[
\text{Metropolis} = e^{-(\text{ESS}' - \text{ESS})/T}
\]

is estimated and compared to Metropolis’, a uniformly distributed random number ranging from [0,1]. If Metropolis is greater than Metropolis’, the new structure is accepted, \( \{ P_s \} \) is updated to \( \{ P_s' \} \), and the algorithm moves uphill. Otherwise, \( \{ P_s' \} \) is rejected and the search for alternative regional grouping schemes minimizing unexplained variance in (1) continues.

From eq. (1), obviously two factors decrease the likelihood of an uphill move: lower ‘temperature’ \( (T) \) and larger differences in the function’s value. After several iterations, the temperature is reduced in steps and the annealing process continues. As temperature is lowered, large moves uphill are discouraged and the algorithm favors smaller refinements leading toward the global minimum. The annealing schedule, that is, the initial temperature and the size of stepwise decreases, is ad hoc and requires experimentation. Successful annealing depends on the schedule and size of perturbations to the system considered at each iteration. The smaller the extent of a perturbation, the more likely the search will efficiently find the
global minimum. The random choice of the initial regional group scheme will also influence the efficiency of the annealing process. The algorithm stops when some preset criterion is met. In general, the algorithm finishes during the final step in cooling after the rate of change in the ESS term fails to meet some preset rate of change related to the CPU speed of the computer doing the various calculations.

We apply these simulated annealing parameters to develop an executable program, which follows the pseudo-code detailed below:

1. Define empirical model (1).
2. Read data into (1).
3. Run a logistic regression with an original regional grouping scheme.
4. Randomly select a regional group. Count the number of countries in it.
5. If there are six or more countries in the group, then randomly choose between changing group based on break up into two groups or randomly swap a country from that group with another group randomly chosen.
6. If there are fewer than six countries in the group, then randomly swap one country from group with another group randomly chosen.
7. Run the logistic regression with new group structure.
8. Compare new ESS with previous ESS and apply Metropolis criterion to accept or reject change in group structure.
9. Repeat steps 3–8 at least 50 times at the given temperature. Stop iterations at given temperature and decrease temperature based on random stopping criterion.
10. Repeat step 9 until final temperature decrease in annealing schedule is accomplished and overall stopping criterion is met.
11. Print final group structure, final logistic regression coefficient estimates and p-values, final pseudo $R^2$ and final ESS.

Once annealing is completed, we are in a position to assess the robustness of the original regional grouping scheme based on three criteria: (1) percentage change in the number of regional groups ($\frac{u_{\text{end}} - u_{\text{beginning}}}{u_{\text{beginning}}}$ where $u$ is the number of regional groups before (beginning) and after (end) annealing); (2) percentage change in overall MNC FDI model explanation (pseudo $R^2_{\text{end}} - \text{pseudo } R^2_{\text{beginning}}$/pseudo $R^2_{\text{beginning}}$ where pseudo $R^2$ is the coefficient of variation before and after annealing and (3) percentage change in MNC FDI model coefficients ($\frac{w_{\text{beginning}} - w_{\text{end}}}{w_{\text{end}}}$ where the difference in $w$ is the number of non-group terms retaining the original coefficient sign and significance after annealing). We can multiply each of these three measures by 100 to obtain percentages of change. A regional grouping
scheme is less (more) robust ex post to the extent that each of these three percentages exceeds (verges on) 0%.

**ILLUSTRATION OF OUR EX POST APPROACH**

*Data Sources, Sampling and Empirical Model for Illustration*

We illustrate this ex post approach to evaluating different regional grouping schemes based on US MNC data used in Flores and Aguilera (2007). They examine the country location of 100 largest US MNCs in 1980 and 2000 based on total sales. Consistent with (1) above, the dependent variable, MNC Subsidiary, is a 0-1 variable taking the value of 1 if the MNC has a subsidiary in the country in the year of observation. These data are obtained from the Directory of American Firms Operating in Foreign Countries (Angel, 1991, 2001), which includes all major US firms’ investments abroad. US MNC investment abroad is where “American firms have a substantial direct capital investment and have been identified by the parent firm as a wholly or partially owned subsidiary, affiliate or branch. Franchises and non-commercial enterprises or institutions, such as hospitals, schools, etc., financed or operated by American philanthropic or religious organizations are not included.” (Angel, 2001, p. i) This operationalization of US foreign location choice allows us to address, at least partially, some of the criticisms of drawing on sales as an overarching measure to capture MNC activities overseas (Clark & Knowles, 2003; Clark et al., 2004; Dunning, Fujita, & Yakova, 2007). US firms in the sample cover 27 different two-digit SIC industry code from oil and gas exploration to pharmaceuticals manufacturing. The US MNCs in this sample have on average substantial direct capital investment on 22.9 countries in 1980 and 28.9 countries in 2000. The total number of substantial foreign capital investments for the 100 MNCs is 2,288 and 2,891 in 1980 and 2000, respectively, an increase of 26% over 20 years.

Again consistent with (1) we define several country-related, MNC-related and time- (year-)related variables. We include in (1) 10 country terms (with expected sign): (1) *Country Wealth* (+), which we operationalize as Gross Domestic Product in billions of current US dollars measures affluence in each year; (2) *Country Size* (+), which we operationalize as the total number of inhabitants in millions; (3) *Country Physical Infrastructure* (+), which we operationalize as the total number of phone lines per thousand inhabitants;
(4) New Country (−), which we operationalize as a 0-1 term where 1 equals a country that did not exist in 1980; (5) Country Political Institutions (+), which we operationalize as a 0-1 term where 1 equals a country judged as democratic; (6) Country Legal System (+), which we operationalize as a 0-1 term where 1 equals a country with an Anglo-American common law tradition; (7) Country Language (+), which we operationalize as a 0-1 term where 1 equals a country where English is an official language; (8) Country Geographic Distance (−), which we operationalize as distance in thousands of miles, between Washington, DC and the capital of each country; (9) Cultural Distance (−), which we operationalize based on Kogut and Singh’s (1988) and (10) Economic Development (+), which we operationalize as a 0-1 term where 1 equals an OECD member country. Data for these terms come from the World Bank’s World Development Indicators (Country Size, Country Physical Infrastructure, New Country), the CIA FactBook (Country Political Institutions, Country Language, Country Economic Development), Reynolds and Flores (1989) (Country Legal System), Great Circle Distances Between Capital Cities (Eden, 2006) (Country Geographic Distance) and International Institute of Culture (Country Cultural Distance).

Again consistent with (1) we include two firm (MNC) terms: (1) Firm Size (+), which we operationalize as the total number of employees and (2) Firm Performance (+), which we operationalize as the total return to investors in the previous 10 years. Data for these variables come from the UN Center for Transnational Corporations (UNCTAD, 2005). Finally, we include a 0-1 Year (−) dummy that equals 1 when year is 1980. We have complete data for foreign investments by 100 US MNCs operating in 105 countries in 1980 and 2000, a total of 19,635 observations for foreign investments made by this group of 100 US firms in 105 countries, total of 19,635 MNC country year observations. Descriptive statistics and pair-wise correlations for this sample are reported in Table 1.

Regional Grouping Schemes and Annealing Schedule for Ex Post Evaluation

The logistic regression model for estimating the likelihood of US MNC investment in various foreign countries, we add regional dummies linked to two regional grouping schemes: (1) four regional dummies corresponding to a five-region grouping scheme based on the continental membership of countries (America (Canada and Latin America/Caribbean), South America, Europe, Africa and Asia) and (2) six regional dummies
Table 1. Descriptive Statistics for Key Variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>s.d.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Capital investment</td>
<td>0.18</td>
<td>0.38</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Firm size</td>
<td>65 e+3</td>
<td>82 e+3</td>
<td>0.10***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Firm performance</td>
<td>11.2</td>
<td>9.6</td>
<td>0.01</td>
<td>-0.02**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Country language</td>
<td>0.34</td>
<td>0.47</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. New country</td>
<td>0.08</td>
<td>0.27</td>
<td>-0.03***</td>
<td>-0.00</td>
<td>0.09***</td>
<td>-0.21***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Country wealth</td>
<td>102.0</td>
<td>326.1</td>
<td>0.33***</td>
<td>-0.00</td>
<td>0.04***</td>
<td>-0.08***</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7. Country population</td>
<td>33.2</td>
<td>119.2</td>
<td>0.12***</td>
<td>-0.00</td>
<td>0.01*</td>
<td>0.00</td>
<td>-0.03***</td>
<td>0.28***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Country physical infrastructure</td>
<td>147.0</td>
<td>182.6</td>
<td>0.32***</td>
<td>-0.00</td>
<td>0.10***</td>
<td>-0.03***</td>
<td>0.18***</td>
<td>0.39***</td>
<td>-0.06***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Country political institutions</td>
<td>0.54</td>
<td>0.50</td>
<td>0.21***</td>
<td>-0.00</td>
<td>0.12***</td>
<td>0.13***</td>
<td>0.13***</td>
<td>0.17***</td>
<td>0.01</td>
<td>0.39***</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Country legal institutions</td>
<td>0.34</td>
<td>0.47</td>
<td>0.12***</td>
<td>-0.00</td>
<td>0.02**</td>
<td>0.61***</td>
<td>-0.13***</td>
<td>0.08***</td>
<td>0.04***</td>
<td>0.07***</td>
<td>0.28***</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>11. Country geographic distance</td>
<td>5.53</td>
<td>2.16</td>
<td>-0.08***</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.26***</td>
<td>-0.06***</td>
<td>-0.05**</td>
<td>0.11***</td>
<td>-0.26***</td>
<td>-0.16***</td>
<td>0.23***</td>
<td>1.00</td>
</tr>
<tr>
<td>12. Country cultural distance</td>
<td>2.81</td>
<td>1.33</td>
<td>-0.14***</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05***</td>
<td>-0.11***</td>
<td>-0.12***</td>
<td>0.00</td>
<td>-0.35***</td>
<td>-0.04***</td>
<td>-0.10***</td>
<td>-0.04***</td>
</tr>
</tbody>
</table>

* p < 0.05.
** p < 0.01.
*** p < 0.001.
corresponding to a seven-region grouping scheme based on Vaaler and McNamara’s (2004) research on sovereign risk rating around the world (North America-Caribbean, Latin America, Western Europe, Central and Eastern Europe, Africa-Middle East, Asia and Oceania). We omit North America from the continental regional grouping scheme, and omit Western Europe from the regional grouping scheme based on Vaaler and McNamara (2004). Once we estimated our base model for each of the two regional schemes, we submit them for iterative re-estimation and simulated annealing according to the following schedule:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial temperature</td>
<td>100</td>
</tr>
<tr>
<td>Temperature reduction factor</td>
<td>0.98</td>
</tr>
<tr>
<td>Ending temperature</td>
<td>0.05</td>
</tr>
<tr>
<td>Maximum steps</td>
<td>100</td>
</tr>
<tr>
<td>Minimum number of iterations/step</td>
<td>50</td>
</tr>
<tr>
<td>Maximum number of iterations/step</td>
<td>25,000</td>
</tr>
<tr>
<td>permitted</td>
<td></td>
</tr>
<tr>
<td>Actual total number of iterations</td>
<td>16,445 (continental), 17,172 (Vaaler &amp; McNamara)</td>
</tr>
<tr>
<td>Running time on workstation computer</td>
<td>8 h</td>
</tr>
</tbody>
</table>

Results before and after Annealing

Results from initial logistic estimation based on both regional grouping schemes are presented in Tables 2-4. Case A results follow from the continental regional scheme while Case B results follow from the Vaaler and McNamara regional scheme. We first examine results in Table 2, that is, the “Beginning” country, firm and year coefficients for Case A and Case B. These coefficients yield intuitive results. For the Case A continental scheme, 11 of the 13 country, firm and year coefficients have the predicted sign and 10 of the 11 are significant at commonly accepted (10% or better) levels. For the Case B Vaaler and McNamara scheme, 10 of the 13 coefficients have the predicted sign and all 10 are significant at commonly accepted levels. US MNCs are more likely to locate FDI in countries abroad if they are more profitable and larger MNCs, if it is in 2000 rather than 1980, and if the host country has the following characteristics: greater wealth and size, better infrastructure, is not newly independent, has more democratic political institutions, a common law legal system, less cultural distance from the US and a higher level of economic development.
What about the explanatory power of each regional grouping scheme at initial estimation? Here, we see clear contrasts in Tables 3 and 4. With Case A’s continental scheme, we see that two regions, Africa and Asia are significantly less likely to receive US MNC FDI, and we note that these four continental dummies as a whole add significant additional explanation to the logistic regression. On the other hand, with Case B’s Vaaler and McNamara scheme, we find no significant regional dummies at the beginning. Were we to stop here, we might conclude that a simple continental scheme emphasizing geography and physical distance adds significantly and practically to overall explanation of MNC FDI patterns over time.

But submission of these two schemes to iterative refinement and re-estimation based on simulated annealing leads to a different view.
We set minimization of the ESS as our annealing criterion and follow the schedule noted above. This way, the algorithm searches for additional model explanation through iterative refinement of each grouping scheme. We track that search in Fig. 1. Two panels in Fig. 1 illustrate along the x-axis the number of iterations, that is, attempted refinements in grouping scheme, over the 100 temperature steps in the algorithm. Along the y-axis, we note changes in overall model explanation using a pseudo-$R^2$ measure commonly reported with logistic regression.

With Case A’s continental scheme, the annealing schedule results in 16,455 iterations over 100 steps while Case B’s Vaaler and McNamara scheme results in 17,172 over 100 steps. The number of iterations per step ranges from the minimum of 50 to more than 1,000. Either the ESS is reduced or because of stochastic criterion permitting acceptance where ESS

<table>
<thead>
<tr>
<th>Regional Dummies</th>
<th>Case A: Continent Regional Scheme$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beginning</td>
</tr>
<tr>
<td>Africa dummy</td>
<td>0.6476$^b$</td>
</tr>
<tr>
<td>America dummy</td>
<td>1.7423</td>
</tr>
<tr>
<td>Asia dummy</td>
<td>0.5030$^b$</td>
</tr>
<tr>
<td>Europe dummy</td>
<td>-1.0066</td>
</tr>
<tr>
<td>Sub-region Africa 1 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region America 1 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Asia 1 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Oceania dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Europe 1 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region America 2 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Africa 2 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Europe 2 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Asia 2 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Africa 3 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region America 4 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Africa 4 dummy</td>
<td></td>
</tr>
<tr>
<td>Sub-region Asia 3 dummy</td>
<td></td>
</tr>
</tbody>
</table>

See Appendix A for a complete description of regional schemes and the countries.

$^a$See Appendix A for a complete description of regional schemes and the countries.

$^b$p < 0.10.

$^p$< 0.05.

$^*$p < 0.01.

$^{**}p$ < 0.001.
We note the similar patterns of change in both panels of Fig. 1. With Case A’s continental scheme, a seemingly random search for refinements to minimize ESS is rewarded approximately 75% of the way through the annealing schedule. At approximately 11,000 iterations, we start an increase in the pseudo-$R^2$ indicating refinement of group structure yielding greater explanation of variation in the likelihood of US MNCE country FDI.

Table 4. Logistic Regression Coefficients for Case B: Beginning and End of Annealing Process.

<table>
<thead>
<tr>
<th>Regional Dummies</th>
<th>Case B: Vaaler and McNamara (2004) Regional Scheme$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beginning</td>
</tr>
<tr>
<td>Africa-Middle East dummy</td>
<td>−1.3906</td>
</tr>
<tr>
<td>Asia dummy</td>
<td>−1.0863</td>
</tr>
<tr>
<td>Central-Eastern Europe dummy</td>
<td>−1.8068</td>
</tr>
<tr>
<td>Latin America dummy</td>
<td>0.99588</td>
</tr>
<tr>
<td>North American-Caribbean dummy</td>
<td>1.1123</td>
</tr>
<tr>
<td>Oceania dummy</td>
<td>−1.3398</td>
</tr>
<tr>
<td>Sub-region Africa-Middle East 1 dummy</td>
<td>3.4168</td>
</tr>
<tr>
<td>Sub-region Asia 1 dummy</td>
<td>2.1287</td>
</tr>
<tr>
<td>Sub-region Central-Eastern Europe 1 dummy</td>
<td>1.7506</td>
</tr>
<tr>
<td>Sub-region Latin America 1 dummy</td>
<td>4.7575</td>
</tr>
<tr>
<td>Sub-region North America-Caribbean 1 dummy</td>
<td>4.8575</td>
</tr>
<tr>
<td>Sub-region not Considered dummy</td>
<td>1.7539</td>
</tr>
<tr>
<td>Sub-region West Europe 1 dummy</td>
<td>3.386</td>
</tr>
<tr>
<td>Sub-region Central-Eastern Europe 2 dummy</td>
<td>−0.1685*</td>
</tr>
<tr>
<td>Sub-region West Europe 2 dummy</td>
<td>2.5836</td>
</tr>
<tr>
<td>Sub-region Africa-Middle East 2 dummy</td>
<td>0.2197*</td>
</tr>
<tr>
<td>Sub-region Africa-Middle East 3 dummy</td>
<td>2.3684</td>
</tr>
<tr>
<td>Sub-region Africa-Middle East 4 dummy</td>
<td>3.9958</td>
</tr>
<tr>
<td>Sub-region Latin America 2 dummy</td>
<td>3.3994</td>
</tr>
<tr>
<td>Sub-region West Europe 2 dummy</td>
<td>1.5181</td>
</tr>
</tbody>
</table>

$^a$See Appendix A for a complete description of regional schemes and the countries.

$^b p<0.10$.

$^* p<0.05$.

$^{**} p<0.01$.

$^{***} p<0.001$.

is not reduced (i.e., Metropolis criterion), we note more than 10,000 changes in both regional grouping schemes over the entire schedule that took approximately 8 h to implement on a state-of-the-art workstation computer.$^2$
Fig. 1. Logistic Regression’s Pseudo-$R^2$ versus Annealing Iterations. (a) Case A: Continents Regional Scheme, (b) Case B: Vaaler and McNamara (2004) Regional Scheme.
The rate of increase begins to level off at approximately 15,000 iterations, near the final steps in the annealing schedule where stochastic jumps based on the Metropolis criterion are quite unlikely. From 11,000 iterations to the end of the annealing schedule at 16,445 iterations, pseudo-$R^2$ increases from 0.32 to 0.38. We observe a $((0.38–0.32)/0.32)$ 18.75% increase in model explanation from the very beginning to the end of simulated annealing.

With Case B’s McNamara and Vaaler scheme, the seemingly random search for refinements to minimize ESS is again rewarded approximately 75% of the way through the annealing schedule. At approximately 11,000 iterations, we again start an increase in the pseudo-$R^2$ indicating refinement of group structure yielding greater explanation of variation in the likelihood of US MNCE country FDI. But only 1,000 iterations or so later, refinements to group structure decrease pseudo-$R^2$ only to see that reversed again in an upward direction at approximately 12,500 iterations. The regional grouping landscape for this scheme is apparently more rugged than in the case of the simpler continental scheme. Even so, we then observe a steady increase in model explanation that begins to level off at approximately 14,000 iterations, near the final steps in the annealing schedule where stochastic jumps based on the Metropolis criterion are less unlikely. From 12,500 iterations to the end of the annealing schedule at 17,172 iterations, pseudo-$R^2$ increases from 0.325 to 0.37. From beginning to end of simulated annealing, we observe a $((0.37–0.325)/0.325)$ 13.8% increase in model explanation.

With annealing completed, we return again to Tables 2–4 for review. We look first at the “End” coefficient estimates in Table 2. Ending firm, country and year coefficients in Table 2 show little change with Case A’s continental scheme. Only 1 of the 13 terms has changed in sign or lost significance at commonly accepted levels. After refinement of the initial group structure, newly independent countries are no longer significantly less likely to have US MNC FDI. This translates into a small $[((11–10)/10) \times 100\%]$ 10% change in key coefficient estimates. Ending firm, country and year coefficients with Case B’s Vaaler and McNamara scheme exhibit only slightly less robustness. After refinement of initial group structure, newly independent countries are no longer significantly less likely nor are English language-speaking countries significantly less likely to have US MNC FDI. This translates into a larger $[((10–8)/8) \times 100\%]$ 25% change in key coefficient estimates.

Tables 3 and 4 report the ending group structures after annealing. In Table 3, Case A’s continental group scheme jumps from 5 (4 dummies) to 14 (13 dummies) sub-continental groups with three new sub-continental
regions significant at commonly accepted levels compared to two at the beginning of the analysis. As a whole, the new group dummies no longer add significantly to overall model explanation at commonly accepted levels. The increase in groups is \[\frac{(14-5)}{5} \times 100\% = 180\%\]. In Table 4, Case B’s Vaaler and McNamara group scheme increases from 7 (6 dummies) to 15 (14 dummies) or an increase of \[\frac{(15-7)}{7} \times 100\% = 143\%\]. Two of the new sub-group dummies are significant at commonly accepted levels, but all of the dummies as a group are not significantly different from zero.

We pull these results together for side-by-side comparison in Table 5 below:

Case A’s continental scheme exhibits more variation in group structure and model explanation but less change in key firm, country and year coefficients explaining MNC FDI compared to Case B’s Vaaler and McNamara scheme. These results prompt more caution in our earlier provisional assessment that simple continental grouping schemes may be preferred to more detailed schemes incorporating geography and level of economic development as in Vaaler and McNamara. The continental scheme of regional dummies may provide significantly more initial explanation before annealing compared to the Vaaler and McNamara scheme, but the continental scheme may also be more sensitive to change in regional group structure and change in overall model explanation. If, on the other hand, the central research aim is to assess the robustness of key coefficients, then our simulated exercise suggests additional support for use of the simpler continental scheme. Even after refinement, more key coefficients retain their original sign and significance compared to the alternative grouping scheme based on Vaaler and McNamara. No matter the research focus, our simulated annealing exercise sheds helpful ex post

### Table 5. Side-by-Side Summary of Results after Simulated Annealing.

<table>
<thead>
<tr>
<th>Annealing Evaluation Criteria</th>
<th>Grouping Scheme</th>
<th>Case A’s Continental Grouping Scheme (Table A1)</th>
<th>Case B’s Vaaler and McNamara Grouping Scheme (Table A2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in group structure (%)</td>
<td>180</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>Change in key coefficients (%)</td>
<td>10</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Change in model explanation (%)</td>
<td>18.75</td>
<td>13.8</td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION AND CONCLUSIONS

Central Results

The central aim of this paper is methodological. We sought to show how empirical models of MNC FDI combined with simulated annealing, can help us to understand the impact of regional grouping schemes on a core IB research phenomenon and debate. We showed conceptually and then through empirical demonstration how regional grouping schemes grounded in intuition or theory (or both) might be subjected to ex post evaluation through a process of iterative refinement and empirical model re-estimation. We developed the general logic for this ex post evaluation method – identifying the extent of before-and-after change through simulated annealing – and identified three potential dimensions for applying that logic. Our application of this method based on Flores and Aguilera (2007) model and two alternative grouping schemes, yielded helpful insight regarding the robustness of each initial group scheme to modest refinement and extended search in a terrain of alternative sub-group structures with many local minima and maxima. Re-estimation with respectively refined group structure yielded additional insight on the robustness of initial model coefficient estimates and overall model explanation.

We think this ex post method for evaluating regional grouping schemes alone or in comparison represent a valuable complementary tool for researchers engaged in understanding the nature and impact of regions on MNC investment behavior. Our method can contribute to current debates over the regionalizing or globalizing nature of MNC expansion by identifying which regional grouping schemes are less (more) robust to reasonable refinement and thus less (more) reliable as indicators of true MNC expansion paths.

Implications for IB Research and Practice

Going forward, we see many implications for IB research and practice. Our method can complement not only ex ante groups defined by geography and/or economic development levels as in this paper, but also across any number of alternative dimensions. For example, we see value in
implementing a series of pair-wise ex post comparisons of grouping schemes: we might consider ex post comparative evaluation of Hofstede’s (2001 [1980]) versus Ronen and Shenkar’s (1985) alternative grouping schemes based on cultural dimensions; we might consider the same for relatively simple law-related grouping schemes proposed by La Porta et al. (1998) versus more complex law-related schemes proposed by Berkowitz et al. (2003). Indeed, we might use simulated annealing to compare any number of culturally, geographically, economically and/or institutionally derived grouping schemes within and across these categories. Our comparative logic and measurable dimensions are sufficiently generic to permit this sort of study and gain greater insight on the value of alternatively defined schemes and their robustness to reasonable refinement.

We also see value extending such methods to other IB and related management phenomena of interest. The group concept is important to many fundamental issues in strategic management. As Fox and et al. (1997) as well as Short, Ketchen, Palmer, and Hult (2007) have demonstrated, groups of firms within an industry space may have collective qualities determining firm behavior and performance as apparently do groups of countries within a geographic, cultural, economic and/or institutional space. If so, then results from initial estimation of strategic group effects for firms will benefit from ex post iterative refinement and re-estimation based on simulated algorithms and evaluative logics and dimensions similar to those developed in this paper. How soon do changes in group structure occur and how quickly do these refinements affect key coefficients and broader model explanation? Our ex post method of evaluating groups based on simulated annealing can render useful research insight across firms grouped within industries, across countries grouped within regions and other grouping designations important to scholars in the broader management field.

NOTES


2. We wrote the program using C++ language and used a MATLAB logistic regression module combined with a simulated annealing algorithm based on (Press, Teukolski, Vetterling, & Flannery, 1992). Interestingly, the MATLAB logistic regression module proved much more time-consuming to implement than the annealing algorithm on our workstation platform.
ACKNOWLEDGMENT

Please contact Paul M. Vaaler regarding this paper. We thank Arash Mahdian and Reza Etebari for research assistance. We thank Howard Guenther and the UIUC Research Board for financial support. All remaining errors are ours.

REFERENCES


A Simulated Annealing Approach


APPENDIX A. TWO REGIONAL GROUPING SCHEMES USED IN SIMULATED ANNEALING ANALYSES

Table A1. Continental Regional Grouping Scheme.

<table>
<thead>
<tr>
<th>Region</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa (43)</td>
<td>Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Congo, Democratic Republic of Congo, Djibouti, Egypt, Ethiopia, Gabon, Gambia, Ghana, Guinea, Ivory Coast, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Senegal, Seychelles, Sierra Leone, South</td>
</tr>
</tbody>
</table>
**Table A1. (Continued)**

<table>
<thead>
<tr>
<th>Region</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa-Middle East</td>
<td>Algeria, Angola, Azerbaijan, Bahrain, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Congo,</td>
</tr>
<tr>
<td>Americas (25)</td>
<td>Argentina, Bahamas, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Surinam, Trinidad &amp; Tobago, Uruguay, Venezuela</td>
</tr>
<tr>
<td>Asia (37)</td>
<td>Azerbaijan, Bahrain, Bangladesh, Brunei, Cambodia, China (PRC), Cyprus, Hong Kong, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Kuwait, Lebanon, Macao, Malaysia, Oman, Pakistan, Philippines, Qatar, Saudi Arabia, Singapore, South Korea, Sri Lanka, Syria, Taiwan (ROC), Thailand, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Vietnam, Yemen</td>
</tr>
<tr>
<td>Europe (37)</td>
<td>Albania, Austria, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Serbia &amp; Montenegro, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom</td>
</tr>
<tr>
<td>Oceania (5)</td>
<td>Australia, Fiji, New Caledonia, New Zealand, Papua New Guinea</td>
</tr>
</tbody>
</table>

Source: [http://unstats.un.org/unsd/methods/m49/m49regin.htm](http://unstats.un.org/unsd/methods/m49/m49regin.htm)

**Table A2.** Vaaler and McNamara (2004) Regional Grouping Scheme.

<table>
<thead>
<tr>
<th>Region</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa-Middle East (55)</td>
<td>Algeria, Angola, Azerbaijan, Bahrain, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Congo,</td>
</tr>
<tr>
<td>Region</td>
<td>Countries</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Democratic Republic of Congo, Djibouti, Egypt, Ethiopia, Gabon, Gambia, Ghana, Guinea, Iran, Iraq, Israel, Ivory Coast, Kenya, Kuwait, Lebanon, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Pakistan, Qatar, Saudi Arabia, Senegal, Seychelles, Sierra Leone, South Africa, Sudan, Swaziland, Tanzania, Tunisia, Uganda, United Arab Emirates, Uzbekistan, Zambia, Zimbabwe</td>
<td></td>
</tr>
<tr>
<td>Asia (25)</td>
<td>Bangladesh, Brunei, Cambodia, China (PRC), Cyprus, Hong Kong, India, Indonesia, Japan, Jordan, Kazakhstan, Macao, Malaysia, Oman, Philippines, Singapore, South Korea, Sri Lanka, Syria, Taiwan (ROC), Thailand, Turkey, Turkmenistan, Vietnam, Yemen</td>
</tr>
<tr>
<td>Central-Eastern Europe (19)</td>
<td>Albania, Belarus, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Malta, Poland, Romania, Russian Federation, Serbia &amp; Montenegro, Slovakja, Slovenia, Ukraine</td>
</tr>
<tr>
<td>Latin America (18)</td>
<td>Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Guyana, Honduras, Nicaragua, Panama, Paraguay, Peru, Surinam, Uruguay, Venezuela</td>
</tr>
<tr>
<td>North America-Caribbean (7)</td>
<td>Bahamas, Canada, Dominican Republic, Haiti, Jamaica, Mexico, Trinidad &amp; Tobago</td>
</tr>
<tr>
<td>Western Europe (18)</td>
<td>Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom</td>
</tr>
<tr>
<td>Not considered (5)</td>
<td>Australia, Fiji, New Caledonia, New Zealand, Papua New Guinea</td>
</tr>
</tbody>
</table>
ADDITIONAL EXEMPLARY CONTRIBUTIONS
This page intentionally left blank
APPLYING ADVANCED PANEL METHODS TO STRATEGIC MANAGEMENT RESEARCH: A TUTORIAL

Peter Hom and Katalin Takacs Haynes

ABSTRACT

This chapter describes how to use popular software programs (Hierarchical Linear Modeling, LISREL) to analyze multiwave panel data. We review prevailing methods for panel data analyzes in strategic management research and identify their limitations. Then, we explain how multilevel and latent growth modeling provide more rigorous methodologies for studying dynamic phenomena. We present an example illustrating how firm performance can initiate temporal change in the human and social capital of members of Board of Directors, using hierarchical linear modeling. With the same data set, we replicate this test with first-order factor latent growth modeling (LGM). Next, we explain how to use second-order factor LGM with panel data on employee cognitions. Finally, we review the relative advantages and disadvantages of these new data-analytical approaches.
Temporal changes in explanatory variables and their dynamic effects on individual or firm outcomes have long fascinated organizational scientists. Toward this end, management scholars have operationalized change in various ways, such as computing difference scores or using repeated measures analysis of variance (RANOVA) (Bergh, 1995; James, Hornick, & Demaree, 1978). Such traditional approaches for analyzing panel data have nonetheless drawn widespread criticisms (Bergh & Fairbank, 2002; Edwards, 1994; Irving & Meyer, 1999). In the wake of such critiques, the dominant methodology for analyzing panel data (repeated waves of measures) has now become autoregressive models that regress an outcome on earlier measures of that outcome and assess the lagged effects of static values of predictor variables (Duncan, Duncan, Strycker, Li, & Alpert, 1999; Williams & Podsakoff, 1989). Going beyond this approach, our chapter reviews alternative, more powerful techniques for analyzing multiwave data that avoids methodological limitations of conventional methods to directly gauge change. In the section that follows, we review common panel-analytic methodologies strategy researchers employ to study change. Then, we describe how two new analytical techniques can offset shortcomings of common methodologies to furnish more valid change assessments and stimulate more scholarly inquiry into dynamic phenomena in the organizational sciences.

LIMITATIONS OF CURRENT APPROACHES FOR ASSESSING CHANGE

Early explorations of how the trajectory of change in a causal variable impacts outcome often computed difference scores between measures of this variable taken on two occasions (e.g., Rusbult & Farrell, 1983). A growing number of critics, however, maintain that difference scores are unreliable, correlate negatively with pretest scores, cannot explain unique variance beyond their components, and reflect regression-toward-the-mean artifacts (Bergh & Fairbank, 2002; Edwards, 1994, 2002; Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991; Irving & Meyer, 1994). Similar shortcomings plague “residualized gain scores” that are computed by predicting an outcome from an earlier assessment of this outcome and then using the residual score (what is not accounted for by the lagged predictor) as a criterion for substantive predictor variables (Irving & Meyer, 1999). Unfortunately, Irving and Meyer (1999) demonstrated that gain scores are essentially weighted difference scores.
Other management scholars have relied on RANOVA to demonstrate how changes in a causal determinant may influence some outcome measure. For example, turnover researchers compare how employees who remain and those who eventually quit differ in mean antecedent levels over time (Hom & Griffeth, 1991; Youngblood, Mobley, & Meglino, 1983). Even so, this method emphasizes between-group change and treats “within-group individual differences in intraindividual change strictly as error” (Francis et al., 1991, p. 59). More than this, RANOVA assumes constant variance across time observations and constant covariances between time observations (Chan, 1998; Cohen, Cohen, West, & Aiken, 2003). Yet this “sphericity” assumption is routinely violated according to Bergh’s (1995) review of 86 repeated measures studies. Because variances of repeated measurements are rarely equal over time, Bergh concluded that Type I errors often inflate statistical tests. Similarly, Chan (1998) argued that sphericity is generally untenable because unequal variances across time observations are substantively meaningful due to systematic interindividual differences in intraindividual changes. Pervasive condemnation of these techniques has limited their use for panel data analyses.

Instead, present-day panel explorations scrutinize how lagged effects of determinants affect outcomes with autoregressive (AR) regression or panel models (Curran, Stice, & Chassin, 1997; Duncan et al., 1999; Kessler & Greenberg, 1981). Measuring outcomes and predictors on two or more occasions, organizational researchers regress outcomes onto earlier predictor measures as well as previous outcome measures. For example, Farkas and Tetrick (1989) surveyed naval personnel on three occasions to evaluate how job satisfaction affects their re-enlistment decisions. They estimated the lagged effect of satisfaction on subsequent re-enlistment decisions, controlling for autoregressive effect of prior reenlistment decisions.

**Autoregressive Panel Model**

![Autoregressive Panel Model Diagram](image-url)
Williams and Podsakoff (1989) later popularized latent variables versions of autoregressive panel models, known as AR structural equations models (SEM) (cf. Dwyer, 1983). Such AR SEM models assess lagged and autoregressive effects of latent variables. Considering an example from Hom and Griffeth (1995), organizational researchers would collect multiple measures of job satisfaction (antecedent) and organizational commitment (consequent) on two or more occasions. Using SEM software, investigators estimate a measurement submodel (that specify various indicators for each latent construct) as well as a structural submodel (that specify causal pathways between latent constructs). In this latter model, they would estimate autoregressive (e.g., effect of Time-1 satisfaction onto Time-2 satisfaction) and cross-lagged (e.g., Time-1 satisfaction onto Time-2 commitment) effects. AR SEM models are superior to AR panel models because they control for random measurement errors as well as autocorrelated errors (e.g., correlation between error terms for Time-1 and Time-2 affect indicators). Importantly, this approach can evaluate alternative “spurious” models, which posit confounding factor that might account for observed cross-lagged effects (cf. Dwyer, 1983).

Table 1 summarizes recent strategy research and identifies commonly deployed panel-data analytical techniques. Cross-lagged panel analyses,
Table 1. Sample List of Strategic Management Articles Published between 2002 and 2007 Examining Temporal Change.

<table>
<thead>
<tr>
<th>Author(s) (Year)</th>
<th>Journal</th>
<th>Topic</th>
<th>Data Analysis Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belke and Heine (2007)</td>
<td>Empirica</td>
<td>Regional business cycles in Europe</td>
<td>Pooled regressions relating moving correlations of residuals</td>
</tr>
<tr>
<td>Goranova, Alessandri, Brandes, and Dharwadhar (2006)</td>
<td>SMJ</td>
<td>Managerial ownership and diversification</td>
<td>Bergh &amp; Fairbank’s component score analysis</td>
</tr>
<tr>
<td>Zhang (2006)</td>
<td>SMJ</td>
<td>Impact of separate COO/president on strategic change and CEO dismissal</td>
<td>Generalized estimating equation to account for the serial correlation of pooled sample</td>
</tr>
<tr>
<td>Wadhwa and Kotha (2006)</td>
<td>AMJ</td>
<td>Knowledge creation</td>
<td>Panel data analysis</td>
</tr>
<tr>
<td>Rothaermel, Hitt, and Jobe (2006)</td>
<td>SMJ</td>
<td>Effects of vertical integration and outsourcing on product portfolio, success and firm performance</td>
<td>GLS with panel data</td>
</tr>
<tr>
<td>Miller (2006)</td>
<td>SMJ</td>
<td>Technological diversity, related diversification and performance</td>
<td>Cross-sectional and panel data</td>
</tr>
<tr>
<td>Kor (2006)</td>
<td>SMJ</td>
<td>Effects of TMT and board composition on R&amp;D investment</td>
<td>Random effects GLS</td>
</tr>
<tr>
<td>Henderson, Danny Miller, and Hambrick (2006)</td>
<td>SMJ</td>
<td>Industry dynamism, CEO tenure and company performance</td>
<td>Natural log of absolute differences</td>
</tr>
<tr>
<td>Fiss (2006)</td>
<td>SMJ</td>
<td>Social influence effects and compensation</td>
<td>Fixed effects pooled time series regression</td>
</tr>
<tr>
<td>David, Yoshikawa, Chari, and Rasheed (2006)</td>
<td>SMJ</td>
<td>Investments in Japanese corporations</td>
<td>Aranello-Bond method to analyze autoregressive distributed lag models of panel data</td>
</tr>
<tr>
<td>Wiggins and Ruefli (2005)</td>
<td>SMJ</td>
<td>Hypercompetition</td>
<td>Outlier analysis using stratification; event history analysis</td>
</tr>
<tr>
<td>Siegel and Hambrick (2005)</td>
<td>Organization Science</td>
<td>Harmful effects of pay disparities on performance</td>
<td>Cross-lagged panel</td>
</tr>
<tr>
<td>Author(s) (Year)</td>
<td>Journal</td>
<td>Topic</td>
<td>Data Analysis Method</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Coombs and Gilley (2005)</td>
<td>SMJ</td>
<td>Stackholder management and CEO compensation</td>
<td>Random effects panel data analysis</td>
</tr>
<tr>
<td>Boyd, Finkelstein, and Gove (2005)</td>
<td>SMJ</td>
<td>Particularism and universalism and the strategy paradigm</td>
<td>Structural equation model using LISREL</td>
</tr>
<tr>
<td>McGahan and Porter (2005)</td>
<td>SMJ</td>
<td>Comments on industry, corp. and business segment effects and business performance</td>
<td>Variance component analysis vs. non-parametric and autoregressive approaches</td>
</tr>
<tr>
<td>Bansal (2005)</td>
<td>SMJ</td>
<td>Corporate sustainable development</td>
<td>Time series cross-sectional data analysis</td>
</tr>
<tr>
<td>Yin and Zajac (2004)</td>
<td>SMJ</td>
<td>Strategy/structure fit in franchising</td>
<td>Random effects panel logit and probit models; random effects pooled cross-sectional time-series regression</td>
</tr>
<tr>
<td>Sanders and Boivie (2004)</td>
<td>SMJ</td>
<td>Valuation of new firms in uncertain markets</td>
<td>Panel autoregressive least squares regression</td>
</tr>
<tr>
<td>Vaaler and McNamara (2004)</td>
<td>Organization Science</td>
<td>Crisis and competition in expert organizational decision making</td>
<td>Cross-lagged panel</td>
</tr>
<tr>
<td>Miller (2004)</td>
<td>SMJ</td>
<td>Technological resources and performance effects of diversification</td>
<td>Wilcoxon signed-rank test</td>
</tr>
<tr>
<td>Jensen and Zajac (2004)</td>
<td>SMJ</td>
<td>Demographic preferences and structural position in corporate elites</td>
<td>Random effects logistic regression using pooled time series</td>
</tr>
<tr>
<td>Nair and Filer (2003)</td>
<td>SMJ</td>
<td>Cointegration of firms within groups in the Japanese steel industry</td>
<td>Cointegration of nonstationary series</td>
</tr>
<tr>
<td>McNamara, Vaaler, and Devers (2003)</td>
<td>SMJ</td>
<td>Hypercompetition</td>
<td>Autoregressive analysis</td>
</tr>
<tr>
<td>Zahra and Nielsen (2002)</td>
<td>SMJ</td>
<td>Sources of capabilities, integration and technology commercialization</td>
<td>Hierarchical multiple regression</td>
</tr>
<tr>
<td>Vermeulen and Barkema (2002)</td>
<td>SMJ</td>
<td>Process dependence in building a profitable MNC</td>
<td>OLS fixed effects models</td>
</tr>
<tr>
<td>Song (2002)</td>
<td>SMJ</td>
<td>Sequential FDI on Japanese firms in East Asia</td>
<td>Random effects logit regression</td>
</tr>
<tr>
<td>Carpenter (2002)</td>
<td>SMJ</td>
<td>TMT heterogeneity and firm performance</td>
<td>Cross-lagged panel</td>
</tr>
</tbody>
</table>
autoregressive OLS regression, and time-series regression are common. Despite popularity and value for examining lagged causal effects, these methods overlook how changes in predictor values per se impact outcome measures. As McArdle and Bell (2000) noted, "the standard autoregressive regression does not explicitly provide parameters for representing changes in the means over time" (p. 73). AR regression or causal models merely assess the time-lagged effects of predictors' static values (collected on earlier occasions) rather than their dynamic effects (cf. Harrison, Virick, & William, 1996; Sturman & Trevor, 2001). Indeed, some panel researchers implicitly rely on "differences between subjects as a proxy for what happens when a variable changes within ... a given set of subjects" (Cohen et al., 2003, p. 569). That is, they infer change from between-person (or between-firm) differences in measured predictors rather than assess change occurring within a subject (Cohen et al., 2003; Dwyer, 1983).

In what follows, we describe two emerging approaches for direct assessment of change and correlates of change, including changes in correlates. These approaches are known as multilevel modeling (Raudenbush & Bryk, 2002) and latent growth modeling (LGM, Chan, 1998). We explain how to use standard statistical software packages to apply these methodologies for estimating change.

ADVANCED TECHNIQUES FOR ASSESSING CHANGE

Hierarchical Linear Modeling

Multilevel modeling approaches can investigate temporal dynamics directly, using Raudenbush and Bryk’s (2002) hierarchical linear modeling (HLM) program (Deadrick, Bennett, & Russell, 1997; Kammeyer-Mueller, Wanberg, Glomb, & Ahlburg, 2005; Short, Ketchen, Bennett, & du Toit, 2006; Sturman & Trevor, 2001). This approach treats repeated assessments of a variable (e.g., board heterogeneity) as an outcome of the timing of their measurement (i.e., level-1 predictor) and other firm characteristics (e.g., firm size, firm performance) as predictors (i.e., level-2 predictors) of temporal change in that variable. In our example, we collected annual measures of the heterogeneity of members on a firm’s board of directors (by computing Blau indices of functional and occupational heterogeneity and summing both indices). To gauge change for board heterogeneity, we would regress the Blau index onto time,
generating an intercept term (representing board heterogeneity at Time-1 – or its initial status) and slope term (representing “rate of change per unit time,” or amount of change in board heterogeneity per year). Thus, change – otherwise known as the “growth trajectory” (a term borrowed from developmental psychology which pioneered advanced panel-analytic techniques) is assessed by the slope parameter. In this company, board heterogeneity is rising over time according to the positive slope parameter (.033).

Multilevel models require three or more assessments of a variable over time. Waves of assessment can be equally spaced (common but not required) or unequally spaced (to better capture rapid changes during some occasions; Singer & Willett, 2003). Impressively, multilevel analyses do not require time-structured data (where each subject is measured at the same time). Data collection schedules can vary across firms. This might be more realistic for organizational studies using panel surveys as respondents rarely submit surveys at the same time (cf. Hom & Griffeth, 1991), especially if the time metric uses finer time units, such as days or weeks. Moreover, multilevel analyses allow for firms to have different
waves (or unbalanced data; Singer & Willett, 2003), including fewer than three waves for some firms. Unfortunately, researchers must delete subjects missing data when testing autoregressive models, compounding the typical subject attrition in panel survey research (Goodman & Blum, 1996).

Importantly, multilevel analyses assess interfirm differences in change. Companies might exhibit different growth trajectories. In our example, another firm shows a negative slope of change for board heterogeneity as well as different initial status. Multilevel analyses permit inclusion of other (level-2) predictors that might account for such between-firm variability in change trajectories.

While more flexible than traditional panel-data analytic approaches, multilevel modeling continues to impose similar conditions, such as the following: (1) the same instrument for assessing a variable is used repeatedly; (2) the measurement has equal validity across all measurement occasions; and (3) scores should not be standardized on each occasion (which removes mean change in the variable over time) (Singer & Willett, 2003).
We use an illustration with actual firm data, examining how changes in board heterogeneity in 52 firms over time may depend on past firm performance. Firm performance data (level-2 predictors) were collected from the year when the firm went public and successive (annual) data on board heterogeneity were collected for up to 13 years later. Firms started at different years and do not yield the same number of measures given differences in firm “age.” Our time variable is clocked from the year after the firm went public and ranges from 1 to 13 corresponding to successive measurement occasions.

Multilevel models simultaneously assess within-firm and between-firm change rather than estimate a separate regression for each firm. These approaches estimate two submodels simultaneously: a level-1 model that describes how each firm changes over time and a level-2 model that describes how changes differ across firms. The level-1 submodel for individual firm change is

$$Y_{ij} = \pi_{0i} + \pi_{1i}(Time_j) + \epsilon_{ij}$$

where \(i\) indexes firms and \(j\) indexes occasions of measurement. The growth parameters are \(\pi_{0i}\) and \(\pi_{1i}\) where

$$\pi_{0i} = \text{the intercept of the change trajectory for company } i$$

and

$$\pi_{1i} = \text{the slope of the change trajectory for company } i$$

or the rate at which company \(i\) changes over time. For better interpretability of the intercept term, the time scale is typically “centered” (e.g., the first measurement is scaled at zero) so that the intercept equals the outcome score (e.g., board heterogeneity) at the first assessment period.

\(\epsilon_{ij}\) represents the residual or deviation of firm \(i\)’s outcome score from the firm trajectory and reflects measurement error and variations in scores not explained by time. Moreover, multilevel models estimate the level-1 residual variance \(\sigma^2\), which represent the scatter of residual scores around firm \(i\)’s change trajectory. A graphic depiction of this level-1 model is shown.
To determine why firms vary in change trajectory, the level-2 model attempts to explain between-firm differences in the growth parameters with (level-2) predictors. (This model assumes firms have the same relationship form between the outcome variable and time, though not necessarily linear.) That is, individual firms will have different level-1 intercept and slope parameters. In essence, the level-2 model regresses individual growth parameters onto level-2 predictors to explain their variation.

In our illustration, we use previous firm performance (ROA, ROE, and market-to-book ratio assessed before the first assessment of board heterogeneity) as level-2 predictors to clarify why firms exhibit different intercepts and slopes for board heterogeneity. Level-2 models estimate the following equation to explain variation in firm intercepts:

\[ \pi_{0i} = \gamma_{00} + \gamma_{01}(\text{ROA}) + \gamma_{02}(\text{ROE}) + \gamma_{03}(\text{MKTBOOK}) + \zeta_{0i} \]

where \( \gamma_{00} \) represents the population average of level-1 intercepts for firms with level-2 predictor scores equal to zero. (For greater interpretability,
level-2 predictors should be “grand-mean centered” by subtracting mean firm performance from each firm’s performance score.) \( \gamma_{01} \) represents the effect of previous ROA on the firm intercept (i.e., how ROA impacts initial board heterogeneity). \( \gamma_{02} \) represents the effect of the earlier year’s ROE on firm intercept (i.e., how ROE impacts initial board heterogeneity). \( \gamma_{03} \) represents the effect of prior market-to-book ratio on the firm intercept. \( \zeta_{0i} \) represents the level-2 residual or between-firm variation in intercept not explained by (level-2) firm performance measures.

To explore why firms have varying slopes, level-2 model also estimates this equation:

\[
p_{1i} = \gamma_{10} + \gamma_{11}(\text{ROA}) + \gamma_{12}(\text{ROE}) + \gamma_{13}(\text{MKTBOOK}) + \zeta_{1i}
\]

where \( \gamma_{10} \) represents the population average of firm slopes for firms having level-2 predictor scores all equal to zero. (If level-2 predictors are grand-mean centered, \( \gamma_{10} \) is the average slope for firms with average performance on all indicators.) \( \gamma_{11} \) represents the effect of past ROA on firm slope (i.e., how ROA affects board heterogeneity’s rate of change). \( \gamma_{12} \) represents the impact of the earlier year’s ROE on the temporal change in board heterogeneity, while \( \gamma_{13} \) represents the influence of past market-to-book ratio on the slope of change. \( \zeta_{1i} \) represents the level-2 residual or between-firm variation in slope that these level-2 predictors cannot explain.

Finally, multilevel analyses estimate variance components. In particular, this method estimates \( \sigma^2_0 \), which is the level-2 residual variance in firm intercepts – or variance of \( \zeta_{0i} \). This quantity represents variation in the level-2 intercept that is not explained by level-2 predictors. Moreover, multilevel analyses estimate \( \sigma^2_1 \), which is the level-2 residual variance in firm slopes (or \( \zeta_{1i} \) variance), capturing variability of firm slopes that remain unaccounted for by level-2 predictors. Further, this method computes \( \sigma^2_{01} \), which is the level-2 residual variance in covariance between firm intercepts and slopes (or covariance between \( \zeta_{0i} \) and \( \zeta_{1i} \)), controlling for level-2 predictors. Finally, HLM yields \( \sigma^2_\epsilon \), which reflects (mean) variability of a firm’s outcome values around a firm’s trajectory (or \( \epsilon_{ij} \) variance). \( \sigma^2_\epsilon \) reflects (residual) variability in outcome scores that time of measurement does not explain.
**HLM Example**

We test this multilevel model for temporal changes in board heterogeneity as a function of firm sales with the HLM software (Raudenbush & Bryk, 2002), given its ease and popularity. We require two data files: a level-1 file comprising the multiple waves of measurement of board heterogeneity (“waves4.sav”) and a level-2 file containing past firm performance measures (“firmdatattime0.sav”). The following chart shows the level-1 file:

This is a firm-period file where multiple waves of assessments are represented as separate rows for each firm. Thus, repeated (five) measurements of board heterogeneity (“BLAUSUM”) for Applied Micro Circuits Corporation are recorded in the first five rows (and grouped by “firmno,” a firm identification variable). This data format varies from traditional formats where each row represents a firm and different variables are positioned on different columns. Our “Time” variable is numbered from 1 (the first wave) to 13 to correspond to 13 waves of measurement, though most firms yield far fewer waves. The first assessment of board heterogeneity (Time=1) begins at different years because firms’ IPO started in different years. The time measure is not centered as the first wave begins with the number “1.” Researchers
can subtract 1 from “Time” to center this variable (which facilitates interpretation of the intercept term) before carrying HLM tests. Alternatively, we use an option in HLM for centering (“group-mean centering,” because our data are unbalanced) the time scale before estimating the multilevel model.

We also require a level-2 data file (“firmdatatime0.sav”), which is shown below:

<table>
<thead>
<tr>
<th>FIRMNO</th>
<th>YEAR</th>
<th>TICKER</th>
<th>ROA</th>
<th>ROE</th>
<th>MKTBOOK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1998</td>
<td>AMCC</td>
<td>.11372341</td>
<td>.14078755</td>
<td>197.68152</td>
</tr>
<tr>
<td>2</td>
<td>2002</td>
<td>AV</td>
<td>-.1709007</td>
<td>-.660.0000</td>
<td>324963.10</td>
</tr>
<tr>
<td>3</td>
<td>2002</td>
<td>AWE</td>
<td>-.0507357</td>
<td>-.117987</td>
<td>1089.2164</td>
</tr>
<tr>
<td>4</td>
<td>1991</td>
<td>AZO</td>
<td>.11019431</td>
<td>.21066683</td>
<td>374.06239</td>
</tr>
<tr>
<td>5</td>
<td>1993</td>
<td>BBBY</td>
<td>.18018737</td>
<td>.28312528</td>
<td>506.73225</td>
</tr>
<tr>
<td>6</td>
<td>1991</td>
<td>BJS</td>
<td>.09192054</td>
<td>.18049325</td>
<td>18.266793</td>
</tr>
<tr>
<td>7</td>
<td>1999</td>
<td>BRCM</td>
<td>.14229578</td>
<td>.16700856</td>
<td>5979.4026</td>
</tr>
<tr>
<td>8</td>
<td>1992</td>
<td>BSX</td>
<td>.18936836</td>
<td>.22742968</td>
<td>858.89979</td>
</tr>
<tr>
<td>9</td>
<td>1997</td>
<td>CJEN</td>
<td>.25254456</td>
<td>.31064348</td>
<td>1857.3248</td>
</tr>
<tr>
<td>10</td>
<td>1994</td>
<td>COF</td>
<td>.03080971</td>
<td>.20074090</td>
<td>147.18372</td>
</tr>
<tr>
<td>11</td>
<td>2002</td>
<td>COL</td>
<td>.09218750</td>
<td>.23910841</td>
<td>772.05761</td>
</tr>
<tr>
<td>12</td>
<td>1997</td>
<td>CPNLQ</td>
<td>.2512680</td>
<td>.14405656</td>
<td>24.947700</td>
</tr>
<tr>
<td>13</td>
<td>1990</td>
<td>CSCQ</td>
<td>.16818066</td>
<td>.20091324</td>
<td>116.19165</td>
</tr>
<tr>
<td>14</td>
<td>1997</td>
<td>CTXQ</td>
<td>.14831999</td>
<td>.21010119</td>
<td>295.14909</td>
</tr>
<tr>
<td>15</td>
<td>1999</td>
<td>CVG</td>
<td>.08676331</td>
<td>.14775669</td>
<td>770.46978</td>
</tr>
<tr>
<td>16</td>
<td>1996</td>
<td>DGX</td>
<td>-.44696955</td>
<td>-.164102</td>
<td>23.322834</td>
</tr>
<tr>
<td>17</td>
<td>2000</td>
<td>DPHIQ</td>
<td>.05334302</td>
<td>.28199681</td>
<td>936.13385</td>
</tr>
<tr>
<td>18</td>
<td>1996</td>
<td>DRI</td>
<td>-.0463553</td>
<td>-.054196</td>
<td>172.60459</td>
</tr>
<tr>
<td>19</td>
<td>1999</td>
<td>EBAY</td>
<td>.01123304</td>
<td>.01270196</td>
<td>2473.5023</td>
</tr>
</tbody>
</table>

This file comprises level-2 predictors: ROA, ROE, and market-to-book ratio (which are lagged predictors for they are collected at Time = 0, before the first assessment of board heterogeneity). This file also includes a firm identifier (“firmno”), a numerical variable that must be present in both files and sorted in ascending order.

We next invoke the HLM program to create a new “multivariate data matrix” (MDM) and select “Stat package input” because data files are SPSS files.
Next, we indicate that our multilevel model has two levels of data (firm-level performance and multiple measures of board heterogeneity by firm) by choosing HLM2:
We next name the new MDM file (HLMBLAU) and identify the type of data files:

We further specify that level-1 data represents multiple measures per firm (or “measures” within persons):
Next, we locate the level-1 firm data ("waves4.sav") for the MDM file:

We open the level-1 file and choose variables for our level-1 model. We choose BDEPTH, BLAUSUM, and TIME (though our level-1 model will use only the last two variables). We also designate FIRMNO as an ID variable (unique identification number for each firm), which will join together level-1 and level-2 data files in the multilevel analysis.
Next, we locate the level-2 file ("firmdatattime0.sav").

After opening this file, we choose variables for the level-2 model (ROA, ROE, and MKTBOOK) as well as designate the ID variable (FIRMNO).
Then we save the MDM template file (which can be later invoked for reanalyses), after naming it “HLMBLAU.”

We press “Make MDM File” next:
After creating the MDM file, output appears momentarily:

This output can be inspected again by pressing “CHECK STATS.”
To proceed, we close this window by hitting “DONE.”

Next, we specify the level-1 model by choosing BLAUSUM (board heterogeneity) as the outcome variable for this equation.
The resulting (incomplete) level-1 model appears next.

To fully specify the level-1 model, we choose the predictor variable, TIME. For greater interpretability, this predictor is “group-mean centered” (rather than grand-mean centered as the number of waves varies across firms).
The final level-1 model next appears:

We then specify the level-2 model.

We first specify the level-2 equation predicting the second-level intercept by selecting ROA as a predictor (using grand-mean centering, which facilitates interpretation of the level-2 intercept).
We next add ROE as another predictor of the level-2 intercept.

We finalize specification of the level-2 intercept equation by adding a third predictor: market-to-book ratio.
The final equation comprising three performance measures predicting the level-2 intercept appears.
Next, we specify the equation predicting the level-2 slope by adding predictors, such as ROA (using grand-mean centering).

We complete this equation by adding the two other level-2 predictors.
We toggle the yellow line where this level-2 equation resides until the residual term appears.

Next, we specify where the output will be placed.
We choose an estimation method next.

We select the “Full Maximum Likelihood” method.
Now, we begin HLM analysis by pressing “Run the Model Shown.”

The HLM analysis will run until the model converges (or stops iterating when the default number of iterations has been reached without converging).
After analyses are completed, the output can be accessed for viewing.

The resulting output reporting level-2 equations is shown below.

The value of the likelihood function at iteration 12 = 2.187840E+002
Circle The outcome variable is BLAUSUM

Final estimation of fixed effects:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>T-ratio</th>
<th>Approx. d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>For INTERCPT, p0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPTZ, B00</td>
<td>1.003153</td>
<td>0.033624</td>
<td>29.835</td>
<td>48</td>
<td>0.000</td>
</tr>
<tr>
<td>ROA, B01</td>
<td>-0.265756</td>
<td>0.300767</td>
<td>-0.884</td>
<td>48</td>
<td>0.382</td>
</tr>
<tr>
<td>ROE, B02</td>
<td>0.007005</td>
<td>0.009803</td>
<td>0.715</td>
<td>48</td>
<td>0.478</td>
</tr>
<tr>
<td>MKTBOOK, B03</td>
<td>0.000013</td>
<td>0.000020</td>
<td>0.657</td>
<td>46</td>
<td>0.514</td>
</tr>
<tr>
<td>For TIME slope, p1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPTZ, B10</td>
<td>0.405246</td>
<td>0.368245</td>
<td>1.100</td>
<td>40</td>
<td>0.277</td>
</tr>
<tr>
<td>ROA, B11</td>
<td>0.044466</td>
<td>0.062375</td>
<td>0.713</td>
<td>48</td>
<td>0.479</td>
</tr>
<tr>
<td>ROE, B12</td>
<td>-0.027611</td>
<td>0.028490</td>
<td>-0.969</td>
<td>48</td>
<td>0.338</td>
</tr>
<tr>
<td>MKTBOOK, B13</td>
<td>0.000006</td>
<td>0.000003</td>
<td>2.090</td>
<td>40</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Circle The outcome variable is BLAUSUM

For the level-2 equation predicting the level-2 intercept, none of the firm performance measures significantly predict this intercept according to \( t \)-tests of statistical significance. In other words, ROA, ROE, and market-to-book
ratio (MKTBOOK) do not have any unique effects on initial board heterogeneity (BLAUSUM). The board heterogeneity intercept is statistically significant from zero: $t = 29.835, p < .05$. At Time-1, the initial level of board heterogeneity for firms having average ROA, ROE, and MKTBOOK (i.e. all centered predictor scores = 0) is 1.00.

For the level-2 equation predicting the level-2 slope, only market-to-book ratio significantly predicted the slope of change in board heterogeneity: $t = 2.09, p < .05$. Thus, this firm performance measure predicts a firm’s rate of change in board heterogeneity. Because this predictor coefficient is positive, firms whose past market-to-book ratios are high develop more heterogeneous boards over time compared with firms whose prior market-to-book ratios are low. To further interpret this finding, we divide firms into those with high market-to-book ratios and those with low ratios and compare their rates of change. The first figure shows firms with high past performance: the rate of ascent of board heterogeneity increases considerably over time. The second figure shows the trajectory for firms with low past performance. For these firms, ascending board heterogeneity trajectory is less steep.
Another set of findings worth interpreting is the significance of the residual variances, reported below.

The residual variance (or $\sigma_0^2$) for the prediction of level-2 intercept is significant according to a $\chi^2$ test (1,707.08, $p < .05$). Thus, the three firm performance measures do not fully explain between-firm differences in initial board heterogeneity; significant residual variance is left unaccounted for. Other performance measures (not considered in this model) thus have the potential to better explain this residual between-firm variance. Similarly, performance measures do not fully account for between-firm variation in slope of change as residual variance ($\sigma_1^2$) in slopes remains significant according to the $\chi^2$ test (210.15, $p < .05$).
Other predictors might help further explain the residual slope differences across firms. Elsewhere, the output reports the covariance ($\sigma_{01}^2$) between intercepts and slopes (controlling for level-2 predictors) to be .00007 (or a .013 correlation).

**Additional Issues for Multilevel Analyses**

Multilevel analyses can examine additional features of multiwave data. The latest multilevel softwares (e.g., HMLM in HLM program) permit various tests of the error variance-covariance structure ($\varepsilon_{ij}$), such as the presence of autocorrelated errors (errors of measurement of the outcome variable correlate over time) and homoscedasticity of error variances. We tested a model that presumes an uncorrelated and homogeneous error structure, which is “rarely tenable in longitudinal research” (Bliese & Ployhart, 2002, p. 382). Bliese and Ployhart (2002) contend that failure to account for autocorrelation can increase Type I error when estimating the relationship between time and the outcome measure by underestimating standard error and inflating statistical tests (though level-2 parameters are the prime focus of most multilevel analyses). Still, Singer and Willett (2003) noted that “regardless of the error structure chosen, estimates of fixed effects (e.g., $\gamma_{10}$, $\gamma_{11}$, $\gamma_{12}$, and $\gamma_{13}$) are unbiased and may not be affected much by choices made in the stochastic part of the model” (p. 264).

Bliese and Ployhart (2002) recommend testing the nature of the error variance-covariance structure with a random coefficient model that omits level-2 predictors (Raudenbush & Bryk, 2002):

\[
\begin{align*}
\text{Level 1} & \quad Y_{ij} = \pi_{0i} + \pi_{1i} (\text{Time}_j) + \varepsilon_{ij} \\
\text{Level 2} & \quad \pi_{0i} = \gamma_{00} + \zeta_{0i} \\
& \quad \pi_{1i} = \gamma_{10} + \zeta_{1i}
\end{align*}
\]

Using HMLM2 (Raudenbush & Bryk, 2002), we compared three different error structures for our board-heterogeneity model: the unrestricted structure, homogeneous $\sigma^2$, and first-order regressive using the first six waves of Blau indices (for more balanced data than the original 13-wave data file). The unrestricted structure always fit data
best as error variances and covariances are not constrained in any way (Singer & Willett, 2003). The homogeneous sigma-squared constrains $\sigma^2_i$ be the same across the measurement occasions. The first-order regressive has a “band-diagonal” shape common for growth processes, a homogeneous $\sigma^2_i$, error covariances that fall with increasing temporal distance between measurements, and a common error autocorrelation (Raudenbush & Bryk, 2002; Singer & Willett, 2003). The output below shows that the unrestricted error variance-covariance matrix fits data best according to $\chi^2$ tests.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Parameters</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Unrestricted</td>
<td>23</td>
<td>-437.74057</td>
</tr>
<tr>
<td>2. Homogeneous sigma_squared</td>
<td>4</td>
<td>-333.49470</td>
</tr>
<tr>
<td>3. First-order Autoregressive</td>
<td>5</td>
<td>-410.54249</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Comparison</th>
<th>Chi_square</th>
<th>df</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 vs Model 2</td>
<td>104.24587</td>
<td>19</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 1 vs Model 3</td>
<td>27.19808</td>
<td>18</td>
<td>0.075</td>
</tr>
<tr>
<td>Model 2 vs Model 3</td>
<td>77.04779</td>
<td>1</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Given its superior fit, we estimated our multilevel model with an unrestricted error structure. Re-estimated level-2 parameters hardly change and do not alter our original conclusion that the past market-to-book ratio accelerates the change trajectory (cf. Short et al., 2006). The new level-2 parameters are presented below.
Besides testing the error structure, we can also test the form of the trajectory (Bliese & Ployhart, 2002). Our example assumed a linear trajectory accurately portrays how board heterogeneity change over time, but other nonlinear forms of change may be more plausible (e.g., quadratic relationship between board heterogeneity and time). However, more complex relationship demand more waves of assessment and can yield highly correlated level-1 time predictors (increasing multicollinearity; Cohen et al., 2003). As Bliese and Ployhart (2002) point out, “few measures have the reliability to permit an assessment of models greater than a cubic model” (p. 381).

**Latent Growth Modeling (LGM)**

Latent growth modeling (LGM; Chan, 1998; Duncan et al., 1999) offers another way to derive measures of change from multiwave data. Like multilevel analyses, LGM analysis presumes that each firm (or individual) possesses its own change trajectory (and unique intercept and slope) and estimates between-firm variability in change trajectory and how other explanatory variables account for such between-firm differences. We discuss first-order factor (FOF) LGM initially, followed by a discussion of second-order factor (SOF) LGM. LGM requires three (or more) waves of data and time-structured (where every firm is measured at the same time) and balanced

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Standard Coefficient</th>
<th>Approx. Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>For INTRCPT1, P0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B00</td>
<td>1.007710</td>
<td>0.030725</td>
<td>32.798</td>
<td>47</td>
<td>0.000</td>
</tr>
<tr>
<td>ROA, B01</td>
<td>-0.295865</td>
<td>0.269277</td>
<td>-1.099</td>
<td>47</td>
<td>0.278</td>
</tr>
<tr>
<td>ROE, B02</td>
<td>0.008987</td>
<td>0.008774</td>
<td>1.024</td>
<td>47</td>
<td>0.311</td>
</tr>
<tr>
<td>MKTBOOK, B03</td>
<td>0.000017</td>
<td>0.000018</td>
<td>0.958</td>
<td>47</td>
<td>0.343</td>
</tr>
<tr>
<td>For TIME slope, P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B10</td>
<td>0.724936</td>
<td>0.402404</td>
<td>1.802</td>
<td>47</td>
<td>0.078</td>
</tr>
<tr>
<td>ROA, B11</td>
<td>0.023740</td>
<td>0.068358</td>
<td>0.347</td>
<td>47</td>
<td>0.730</td>
</tr>
<tr>
<td>ROE, B12</td>
<td>-0.050130</td>
<td>0.030567</td>
<td>-1.640</td>
<td>47</td>
<td>0.107</td>
</tr>
<tr>
<td>MKTBOOK, B13</td>
<td>0.000009</td>
<td>0.000003</td>
<td>3.547</td>
<td>47</td>
<td>0.001</td>
</tr>
</tbody>
</table>
data (the same number of waves for each firm). LGM applications do not require measurement occasions to be equally spaced, though such schedules are common in panel studies. Though more stringent in its data requirements than HLM (Bliese & Ployhart, 2002), LGM do confer certain methodological advantages worth pursuing (Lance, Vandenberg, & Self, 2000).

LGM is a form of confirmatory factor analysis (CFA) that represents the intercept and slope as latent “growth” factors that underpin repeated measurement of a particular variable. In general, FOF LGM estimates a growth model depicted in this figure.

We use LISREL notation to describe this model and how this model can be assessed with the LISREL software (Jøreskog & Sorbom, 2001). In this growth model, four repeated measures of the same variable ($x_1$ to $x_4$) (taken at equally spaced time intervals) define a variable’s growth factors (i.e., initial status and rate of change). Unlike CFA, indicators from different occasions have fixed (rather than estimated) factor loadings ($\lambda$) on the growth factors in this model. To identify the intercept factor, all indicators have a common fixed factor loading of 1.0 (because an intercept is constant over time). For the sake of simplicity, we posit a linear trajectory of change for this model. To specify a linear slope, the factor loading for the first indicator is fixed at 0, the loading for the second indicator ($x_2$) is fixed at 1, the loading for $x_3$ is fixed at 2, and the $x_3$ loading is set at 3. When plotted as a function of time, these factor loadings fall along a straight line, depicting a linear relationship with measurement occasions (at equally
spaced time intervals). For nonlinear change, the pattern of slope factor loadings would depict a curvilinear function of timing of measurement.

For Linear Change
- Time 1, $\lambda_s = 0$
- Time 2, $\lambda_s = 1$
- Time 3, $\lambda_s = 2$
- Time 4, $\lambda_s = 3$

For Quadratic Change
- Time 1, $\lambda_s = 0$
- Time 2, $\lambda_s = 1$
- Time 3, $\lambda_s = 4$
- Time 4, $\lambda_s = 9$

Each indicator at time $t$ is thus a function of both the intercept ($\xi_i$) and slope ($\xi_s$) factors (as well as measurement error): $x(t) = \lambda_s \xi_s + \lambda_i \xi_i + \delta(t)$. Though not apparent in the figure, growth factors are presumed to vary across firms (or individuals). Rather than estimate firms’ growth factor scores directly, FOF LGM analysis generates summary statistics for growth factors – namely, latent variable means (average slope: $\kappa_s$; average intercept: $\kappa_i$), their variance across firms (slope variance: $\phi_s$; intercept variance: $\phi_i$), and covariance between firms’ intercept and slope ($\phi_{is}$). Importantly, LGM assesses structural relationships between growth factors and external variables, such as effects of the intercept ($\gamma_i$) and slope ($\gamma_s$) factors on an outcome.

**FOF LGM Illustration**

For illustration, we use the first four assessments of board heterogeneity (Time = 0–4, beginning with the year of the IPO) to estimate growth parameters to predict firm performance (Tobin’s $Q$ 2 years after the last assessment of board heterogeneity). Thus, we demonstrate how the intercept
and rate of change for board heterogeneity might predict subsequent firm performance. Unlike the HLM example, all data for a given firm resides on one row with separate columns representing variables. We use LISREL to carry out FOF LGM analysis.

First, we specify the measurement submodel for the latent growth factors by specifying the appropriate pattern of fixed factor loadings for the intercept and slope factors.

**LISREL MATRIX NOTATION**

\[ \Lambda_X = \begin{bmatrix} \text{Indicators} \\
\text{Blau0} & \text{Blau1} & \text{Blau2} & \text{Blau3} \\
1 & 1 & 1 & 1 \\
0 & 1 & 2 & 3 \end{bmatrix} \]

\[ \text{SPECIFY FIXED FACTOR LOADINGS FOR } \Lambda_X \text{ (LAMBDA MATRIX)} \]
Next, we specify the desired properties for the PHI (PH) matrix (covariance matrix for the growth factors) and the THETA-DELTA (TD) Matrix (covariance matrix for the indicator measurement errors).

**DESIZED PH AND TD MATRICES**

**Factor Variance-Covariance Matrix (PHI MATRIX)**

\[
\Phi = \begin{bmatrix}
\phi_{II} & \phi_{IS} \\
\phi_{SI} & \phi_{SS}
\end{bmatrix}
\]

**Measurement Error Variance (THETA-DELTA MATRIX)**

\[
\Theta_\delta = \begin{bmatrix}
\Theta_\delta & 0 & 0 & 0 \\
0 & \Theta_\delta & 0 & 0 \\
0 & 0 & \Theta_\delta & 0 \\
0 & 0 & 0 & \Theta_\delta
\end{bmatrix}
\]

Then we estimate all elements in the KAPPA (KA) vector to estimate means for the intercept and growth factors.

**LATENT VARIABLE MEANS**

**ESTIMATE VECTOR OF MEANS FOR GROWTH FACTORS (KAPPA VECTOR)**

\[
K = \begin{bmatrix}
\kappa_i \\
\kappa_s
\end{bmatrix}
\]

Mean Intercept

Mean Slope

We next specify the measurement submodel for the firm performance measure. Because there is only one indicator, the indicator is synonymous
with the construct. Thus, the LAMBDA-Y factor matrix includes a factor loading of 1.0, while the measurement error (THETA-EPSILON) matrix includes an entry of 0.

\[
\Lambda_Y = \begin{bmatrix} 1 \end{bmatrix} \quad \Theta_\epsilon = \begin{bmatrix} 0 \end{bmatrix}
\]

We next specify the GAMMA (GA) matrix to estimate the impact of the growth factors and PSI (PS) matrix to estimate the residual variance for the firm performance measure.

\[
\Gamma = \begin{bmatrix} \gamma_i & \gamma_s \end{bmatrix}, \quad \Psi = \begin{bmatrix} \psi \end{bmatrix}
\]
After specifying LISREL matrices, we write the LISREL commands.

**LISREL COMMANDS**

- **TITLE COMMAND**
  - FIRST-ORDER FACTOR LATENT GROWTH MODELING
  - PREDICT TOBIN’S Q FROM GROWTH PARAMETERS FOR BOARD HETEROGENEITY (4 OCCASIONS)

- **DA**
  - NI = 5 (FIVE VARIABLES)
  - NO = 47 (NO. OF FIRMS)
  - MA = CM (SPECIFY THAT COVARIANCE MATRIX FOR ANALYSIS)

- **RA** =
  - SPECIFY LOCATION OF DATA FILE

Next, we specify model matrices.

**LISREL COMMANDS**

- **MO** (Model Specification)
  - NX = 4 (NO. OF BLAU INDICES)
  - NK = 2 (NO. OF FACTORS: INTERCEPT & SLOPE FACTORS)
  - NY = 1 (NO. OF PERFORMANCE MEASURES)
  - NE = 1 (NUMBER OF PERFORMANCE FACTORS)
  - PH = SY, FR
    - ESTIMATE ALL ELEMENTS IN GROWTH FACTOR VARIANCE-COVARIANCE MATRIX
  - LX = FU, FI
    - FIX ALL FACTOR LOADINGS IN FACTOR MATRIX FOR BLAU INDICES TO 0
  - TD = DI
    - ESTIMATE ERROR VARIANCES FOR BLAU INDICES
  - KA = FR
    - ESTIMATE GROWTH FACTOR MEANS
      - MEAN INTERCEPT VALUE
      - MEAN SLOPE VALUE

\[
\begin{align*}
\Phi &= \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \\
\Theta &= \begin{bmatrix}
\Theta_4 & 0 & 0 \\
0 & \Theta_5 & 0 \\
0 & 0 & \Theta_6
\end{bmatrix}
\end{align*}
\]
Then, we complete specification of the various LISREL matrices

**LISREL COMMANDS**

- MO (Continued)
  - $LY = FU, FI$ (ROE FACTOR LOADING IS FIXED AT 0)
  - $TE = DI, FI$ (SET ERROR VARIANCE OF ROE MEASURE TO 0)
  - $GA = FU, FR$ (ESTIMATE STRUCTURAL EFFECTS OF INTERCEPT AND SLOPE FACTORS ON ROE)
  - $PS = FR$ (ESTIMATE THE VARIANCE OF ROE DISTURBANCE TERM)

Next, we set factor loadings for the growth factors.

**LISREL COMMANDS**

- $LK$ (INTERCEPT SLOPE)
  (LABEL GROWTH FACTORS)
- $LE$ (ROE)
  (LABEL ROE FACTOR)
- $VA\, LX\ (1,1)\, LX\ (2,1)\, LX\ (3,1)\, LX\ (4,1)$
  - SET ALL FACTOR LOADINGS FOR INTERCEPT FACTOR TO 1
- $VA\, 0\, LX\ (1,2)$
- $VA\, 1\, LX\ (2,2)$
- $VA\ 2\, LX\ (3,2)$
- $VA\ 3\, LX\ (4,2)$
  - SET FACTOR LOADINGS FOR SLOPE FACTOR TO 0, 1, 2, & 3
- $VA\ 1\, LY\ (1,1)$
  - SET ROE FACTOR LOADING TO 1
- OU SC SE TV IT=900
  - REQUEST STANDARDIZED ESTIMATES, STANDARD ERRORS, T-TESTS

$$\Lambda_Y = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Theta_\epsilon = \begin{bmatrix} \gamma_i \\ \gamma_s \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \gamma_i \\ \gamma_s \end{bmatrix}$$

$$\Psi = \begin{bmatrix} \psi \end{bmatrix}$$

$$\Lambda_X = \begin{bmatrix} \Lambda_Y \\ \Lambda_X \end{bmatrix}$$

$$\Lambda_Y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
We next run the LISREL analysis using the command statements below.

The following output appears, showing factor matrices:

**LISREL OUTPUT**

\[ \Lambda_y \]

\[ \begin{align*}
\text{Tobinsq} & \\
1.00 & \\
\end{align*} \]

\[ \Lambda_x \]

\[ \begin{align*}
\text{Intercep} & \quad \text{Slope} \\
\text{Blau0} & \\
1.00 & - - \\
\text{Blau1} & \\
1.00 & 1.00 \\
\text{Blau2} & \\
1.00 & 2.00 \\
\text{Blau3} & \\
1.00 & 3.00 \\
\end{align*} \]
The following shows error variances for measures of board heterogeneity:

**Measurement Error Variances**

\[ \Theta, \delta \]

<table>
<thead>
<tr>
<th></th>
<th>BLAU0</th>
<th>BLAU1</th>
<th>BLAU2</th>
<th>BLAU3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>3.31</td>
<td>3.11</td>
<td>2.71</td>
<td>1.84</td>
<td></td>
</tr>
</tbody>
</table>

Of greater interest are parameter estimates for the growth factors. The PHI matrix reports variances and covariance for growth factors. Importantly, both variances are statistically significant (\(t\)-tests in third row), indicating significant firm variability in initial status of board heterogeneity and rate of change over 4 years. There is no significant covariation between slope and intercept factors. The KAPPA vector shows that average board heterogeneity on the first measurement occasion significantly differs from zero and that there is no significant mean rate of change in board heterogeneity.

**Growth Factors**

**PHI**

<table>
<thead>
<tr>
<th></th>
<th>INTERCEPT</th>
<th>SLOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.12</td>
<td></td>
</tr>
<tr>
<td>SLOPE</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td></td>
<td>-0.39</td>
<td>2.10</td>
</tr>
</tbody>
</table>

**KAPPA**

<table>
<thead>
<tr>
<th></th>
<th>INTERCEPT</th>
<th>SLOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28.66</td>
<td>0.14</td>
</tr>
</tbody>
</table>
Next, we inspect GAMMA parameters. The intercept significantly predicts Tobin’s $Q$; firms with more heterogeneous boards during the first measurement occasion perform more effectively. The slope did not significantly predict Tobin’s $Q$. The PSI matrix indicates that there is significant residual variance in Tobin’s $Q$ that is not explained by growth factors.

**PERFORMANCE PREDICTIONS**

<table>
<thead>
<tr>
<th>GAMMA</th>
<th>INTERCEPT</th>
<th>SLOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOBINSQ</td>
<td>2.80</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(13.43)</td>
</tr>
<tr>
<td></td>
<td>6.72</td>
<td>0.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PSI</th>
<th>TOBINSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.30</td>
</tr>
<tr>
<td></td>
<td>(1.73)</td>
</tr>
<tr>
<td></td>
<td>4.79</td>
</tr>
</tbody>
</table>

Fit statistics indicate that this linear trajectory growth model fits data: NFI = .91 and CFI = .95.

**Goodness of Fit Statistics**

- Degrees of Freedom = 8
- Minimum Fit Function Chi-Square = 16.22 (P = 0.039)

- Normed Fit Index (NFI) = 0.91
- Non-Normed Fit Index (NNFI) = 0.94
- Parsimony Normed Fit Index (PNFI) = 0.73
- Comparative Fit Index (CFI) = 0.95
- Incremental Fit Index (IFI) = 0.95
- Relative Fit Index (RFI) = 0.89
- Critical N (CN) = 57.97
- Root Mean Square Residual (RMR) = 0.035
- Standardized RMR = 0.077
- Goodness of Fit Index (GFI) = 0.90
- Adjusted Goodness of Fit Index (AGFI) = 0.81
- Parsimony Goodness of Fit Index (PGFI) = 0.48
Second-Order Factor Latent Growth Modeling

While LGM does not seem more useful than HLM (especially given its more stringent data requirements), this method offers special methodological benefits when multiple measurements of a variable are available. Given such data, SOF LGM can address one of the thorniest problems in longitudinal data – namely, artifactual shifts in psychometric properties of indicators can confound interpretations of change (Chan, 1998, 2002; Vandenberg & Lance, 2000). In particular, “beta” and “gamma” changes can undermine valid inferences about observed changes in variable means and interrelationships over time – otherwise known as “alpha” change (Chan, 2002). Beta change occurs when a measurement scale is recalibrated over time, such as respondents interpreting a rating scale differently on different occasions. For example, the management department in the W. P. Carey School of Business relies on faculty peer reviews to assess faculty performance. When this practice began, faculty would interpret “standard” performance as a “2” on a seven-point scale and assign some colleagues mean ratings of around 2.0. The department chair, however, notified faculty that higher administration viewed such ratings as grounds for dismissal or removal of tenure. Subsequently, management faculty would rate standard performance as a 4.0, primarily rating from 4 to 7 (essentially truncating the performance scale). That is, raters changed their definition of scale values over time and their frame of reference. Thus, mean improvements in peer ratings would not represent actual performance gains over time. Rather, upward changes in performance ratings would be spurious as raters’ definition of standard performance shifted upwards from a “2.0” to a “4.0.”

Gamma change, where the meaning of a measure changes over time, represents another serious threat to valid inferences about change (Vandenberg & Self, 1993). For example, survey respondents may redefine the construct being assessed by a question. Going back to our example, management faculty at Arizona State University are judging each other’s teaching effectiveness differently over time. Formerly, they defined excellent teaching as simply high student evaluations. Now, they define teaching effectiveness in a more global manner, relying on broader array of evidence of effective teaching, such as difficulty of subject matter and rigid in course assignments. In other words, the underlying construct measured by seven-point teaching rating scale changed over time (formerly reflecting student evaluations and now reflecting a more multidimensional construct). In summary, “when there is gamma or beta change over time, it is not meaningful to … interpret the change pattern over time” (Chan, 2002, p. 335).
To check for these artifacts, Vandenberg and Lance (2000) advocate testing longitudinal measurement invariance before assessing change. According to Lance et al. (2000), “the unambiguous interpretation of change on some measured variable requires the prerequisite assumption of measurement equivalence across measurement occasions.” In essence, Vandenberg and his colleagues recommend longitudinal CFA to test the following type of measurement model.

In this model, a latent construct is measured by three indicators (Ys) on three separate occasions. Unlike typical models tested in CFA, this longitudinal measurement represents each occasion as a common factor (e.g., Time-1) assessed by three indicators (three indicators are needed for model identification). Moreover, this model specifies each indicator as a function of an intercept term ($\tau$) as well as the latent factor and measurement error:

$$y = \tau + \lambda(\text{Latent Factor}) + \varepsilon$$

Evaluating the longitudinal stability of the intercept term is invaluable for valid comparisons of mean-level change in a variable over time. This longitudinal measurement model further allows for covariances ($\theta$) between the same indicator over time to capture autocorrelated errors.

Vandenberg and associates prescribe a series of nested model tests for this longitudinal model (Lance et al., 2000; Vandenberg & Lance, 2000).
First, we test “configural invariance” to evaluate Gamma change by assessing the fit to data of this measurement model (Model 1). Should this model fit data (according to omnibus fit indices, such as the CFI, and significant factor loadings), we conclude that the number of factors (one factor in this example) and pattern of factor loadings (e.g. Y1 loads on Factor 1 but not on Factors 2 or 3) is stable over time. If this model does not fit data, we do not proceed any further with this construct. It would be meaningless to examine its temporal change (Lance et al., 2000).

**Nested Measurement Models**

- **First Test**
  - Full “Configural Invariance”
    - Does Nature Of The Construct Operationalized By Measures Remain Stable Over Time? (Gamma Change)
    - Number of Factors are Identical over Time?
    - Identical Pattern of Factor Loadings?

- **Second Test**
  - Metric Invariance
    - Relations between measures and corresponding factors are stable over time? (Gamma Change)
    - Factor Loadings Are Identical Over Time?
      - \( \lambda = \lambda' \)
      - Partial metric invariance is OK

- **Third Test**
  - Scalar Invariance
    - Measures have the same intercept terms over time? (Beta Change)
      - Usually Assumed to be Zero by Default
    - Partial scalar invariance is OK

If this test supports configural invariance, we proceed with a second test by constraining factor loadings to be identical over time in the longitudinal measurement model. This “metric invariance” test ascertains gamma change by determining if indicators have stable validity – reflect the latent construct with the same precision over time. This second model (Model 2) is a “nested” version of Model 1, derived from the latter by adding equality constraints on the factor loadings. Thus, an indicator that is repeated over time would have the same factor loading on all occasions. We then test the fit of Model 2, which posits all invariant factor loadings. If Model 1 fits data as well as Model 2, then the test demonstrated metric invariance. (The logic
behind nested model testing is to accept a more constrained model if it fits as well as the less constrained model for the sake of parsimony.)

Suppose Model 2 fits worse than Model 1. We might improve the fit of Model 2 by allowing a “few” factor loadings to fluctuate over time (assuming some theoretical justification and only a few indicators are involved; Vandenberg & Lance, 2000). We would then retest this model by relaxing a few equality constraints. If this respecified model fits as well as Model 1, we can still proceed with change assessments. As will be shown later, SOF LGM (unlike other panel data-analytic tools) can accommodate partial measurement invariance (in factor loadings or intercepts) to correct their biasing effects (Lance et al., 2000).

After demonstrating full or partial metric invariance, we carry out our third test to evaluate “scalar invariance,” checking if intercepts are stable over time and thus assessing beta change (Vandenberg & Self, 1993). We thus test a third measurement model by specifying that each indicator’s intercept be constant across occasions (Model 3). This model retains equality constraints imposed on the factor loadings in Model 2. Therefore, Model 3 is a nested version of Model 2. If this model fits as well as Model 2, then we obtain evidence for scalar invariance. If Model 3 yields worse fit, we might also “salvage” this model by allowing a “few” intercepts to vary over time. We then re-estimate Model 3 with fewer intercept constraints. If this revised model (specifying partial scalar invariance) fits, we can proceed with SOF LGM to assess changes in latent variable means (Lance et al., 2000). In summary, researchers can assess change with SOF LGM after demonstrating these preconditions for measurement invariance.

Illustration of Measurement Invariance Tests

For an illustration, we use an example from organizational behavior as the board-capital data has too few cases (firms) to permit tests of longitudinal measurement invariance (or SOF LGM). We use data from Hom, Griffeth, and Sager (2006) who surveyed 290 working students on three occasions (spaced a month apart), asking them to report their thoughts about leaving their current employer (withdrawal cognitions). Students answered three questions about their thoughts of quitting (THINKQ), intention to search for another job (BILOOK), and intention to quit (BIQUIT), using five-point Likert rating scales. Specifically, we test the invariance of these indicators of withdrawal cognitions with the following longitudinal measurement model.
To facilitate invariance testing, we group indicators together from the same measurement period on the data file. Thus, the first three measures of withdrawal cognitions appear in the first three columns, while the next set of measures appear in the next three columns, etc.
We use Y-side LISREL equations to facilitate expression of this longitudinal measurement model. We specify the LAMBDA-Y (LY) matrix – or factor matrix for indicators – on the MO command line in LISREL. We declare a “reference indicator” for each factor by fixing a factor loading to 1 (which establishes the scale of the latent factor; Bryne, 1998).

**SPECIFYING LY MATRIX**

```
MO.......LY = FR

<table>
<thead>
<tr>
<th>LE</th>
<th>WCGTIME1</th>
<th>WCGTIME2</th>
<th>WCGTIME3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA LY</td>
<td>1 0 0</td>
<td>1 0 0</td>
<td>1 0 0</td>
</tr>
<tr>
<td>1 0 0</td>
<td>1 0 0</td>
<td>1 0 0</td>
<td></td>
</tr>
<tr>
<td>1 0 0</td>
<td>1 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 0</td>
<td>1 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 1</td>
<td>0 0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARAMETER MATRIX</th>
<th>THINFQ (ROW 1)</th>
<th>BLOCK1 (ROW 2)</th>
<th>BIEIXT1 (ROW 3)</th>
<th>Theta1 = Y1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME THE THREE OCCASION FACTORS</td>
<td>BIEIXT2 (ROW 4)</td>
<td>THINFQ (ROW 5)</td>
<td>BLOCK2 (ROW 6)</td>
<td>Theta2 = Y2</td>
</tr>
<tr>
<td>WAY TO SPECIFY FACTOR PATTERN MATRIX</td>
<td>BIEIXT3 (ROW 7)</td>
<td>THINFQ (ROW 8)</td>
<td>BLOCK3 (ROW 9)</td>
<td>Theta3 = Y3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WCGTIME1 (COLUMN 1)</th>
<th>WCGTIME2 (COLUMN 2)</th>
<th>WCGTIME3 (COLUMN 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\lambda_{11} = 1</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>\lambda_{21}</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>\lambda_{31}</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>\lambda_{42} = 1</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>\lambda_{52}</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>\lambda_{62}</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>\lambda_{73} = 1</td>
<td>0 0</td>
<td></td>
</tr>
</tbody>
</table>

| VA 1 LY (1 1) LY (4 2) LY (7 3) |
```

We next specify the TE matrix on the LISREL MO command line. For longitudinal measurement models, we generally estimate autocorrelated errors in the TE covariance matrix.

**SPECIFY TE MATRIX**

```
MO..TE=SY,FR

<table>
<thead>
<tr>
<th>PA TE</th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
<th>Y6</th>
<th>Y7</th>
<th>Y8</th>
<th>Y9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>\theta_{12}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1</td>
<td>0</td>
<td>\theta_{23}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 1</td>
<td>0</td>
<td>\theta_{34}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 0 1</td>
<td>0</td>
<td>\theta_{45}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1 0 0 1</td>
<td>0</td>
<td>\theta_{56}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 0 1 0 1</td>
<td>0</td>
<td>\theta_{67}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1 0 0 1 0 1</td>
<td>0</td>
<td>\theta_{78}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 1 0 0 1 0 1</td>
<td>0</td>
<td>\theta_{89}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPECIFY TE AS SYMMETRIC &amp; FREE</th>
</tr>
</thead>
</table>
```

Applying Advanced Panel Methods to Strategic Management Research 243
On the MO line, we also specify the PSI (PS) matrix to estimate variances and covariances among the three occasion factors.

**SPECIFY PS MATRIX**

```
MO...PS=SY,FR
```

- Specify PS as symmetric & free
- Estimate the diagonal entries (factor variance)
- Estimate the factor covariances (off-diagonal entries)
- Symmetric means that estimating one diagonal entry automatically estimates the corresponding diagonal entry.
  
  E.g., estimating $\psi_{21}$ implies estimating $\psi_{12}$ because they are equal ($\psi_{21} = \psi_{12}$)

Finally, we estimate the intercepts (TAU-Y vector) and factor means (ALPHA vector). For indicators chosen as reference indicators (Y1, Y4, and Y7), we set their intercepts to zero. For identification purposes, we thus fix intercepts ($\tau$) of reference indicators to zero, which locates the latent factor mean to the item’s mean.

**SPECIFY TY & AL VECTORS**

```
TY
```

```
AL
```

SET INTERCEPTS FOR REFERENCE INDICATORS TO ZERO

ESTIMATE FACTOR MEANS
Now we run the following LISREL program to first test configural invariance.

```
TEST LONGITUDINAL MEASUREMENT INVARINANCE
FOR MEASURES OF WITHDRAWAL COGNITIONS
Working Students - 3 survey waves
FIRST TEST OF CONFIGURAL INVARINANCE
DA NI=9 NO=290 MA=CM
RA='C:\WITHCOG.PSF'
MO NY=9 NE=3 LY=FR TE=SY,FR PS=SY,FR TY=FR AL=FR
LE
WCGTIME1 WCGTIME2 WCGTIME3
PA LY
1 0 0
1 0 0
1 0 0
0 1 0
0 1 0
0 0 1
0 0 1
0 0 1
FI LY 1 1 LY 4 2 LY 7 3
VA 1 LY 1 1 LY 4 2 LY 7 3
FI TY 1
VA 0.0 TY 1
FI TY 4
VA 0.0 TY 4
FI TY 7
VA 0.0 TY 7
PA TE
* 
1
0 1
0 0 1
1 0 0 1
0 1 0 0 1
0 0 1 0 0 1
1 0 0 1 0 0 1
0 1 0 0 1 0 0 1
0 0 1 0 0 1 0 0 1
OU SS SC SE TV IT=600
```

The following LISREL output results showing the factor loadings. While the first (least-restrictive) measurement model lets factor loadings vary across time, these loadings seem stable (e.g., BIEXIT factor loadings seem similar over time). This is not surprising given the short time span (one month) between assessments.
We next inspect the PSI (PS) matrix (comprising factor variance and covariances).

Next, we can review the THETA-EPILSON (TE) matrix, which comprises error variances and covariances. This matrix reveals
autocorrelated measurement errors – errors for a given indicator are indeed correlated across different time periods.

<table>
<thead>
<tr>
<th>TETA-EPS</th>
<th>THINKQ1</th>
<th>BILOOK1</th>
<th>BIEXIT1</th>
<th>THINKQ2</th>
<th>BILOOK2</th>
<th>BIEXIT2</th>
</tr>
</thead>
<tbody>
<tr>
<td>THINKQ1</td>
<td>0.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>BILOOK1</td>
<td>-</td>
<td>0.49</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>BIEXIT1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.32</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>THINKQ2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.39</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>BILOOK2</td>
<td>-</td>
<td>0.19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>BIEXIT2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.06</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>THINKQ3</td>
<td>0.25</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.27</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td></td>
<td></td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>BILOOK3</td>
<td>-</td>
<td>0.17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>BIEXIT3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

Intercepts for reference indicators are indeed fixed at zero in the TAU-Y (TY) vector, which are shown as dashed lines in the output. Interestingly, factor means in the ALPHA vector indicate that withdrawal cognitions are increasingly over time.

<table>
<thead>
<tr>
<th>TY &amp; AL VECTORS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INTERCEPTS</strong></td>
</tr>
<tr>
<td>TAU-Y</td>
</tr>
<tr>
<td>THINKQ1</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>THINKQ3</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALPHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCGETIME1</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Intercept = 0 for Reference Indicators**
Finally, indices of overall model fit reveal that this measurement model fits data well. In conclusion, this test supports configural invariance for measures of withdrawal cognitions. Thus, we proceed with the next test of metric invariance.

Goodness of Fit Indices

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees of Freedom</td>
<td>15</td>
</tr>
<tr>
<td>Minimum Fit Function Chi-Square</td>
<td>14.20 (P = 0.51)</td>
</tr>
<tr>
<td>Normal Theory Weighted Least Squares Chi-Square</td>
<td>14.07 (P = 0.52)</td>
</tr>
<tr>
<td>Estimated Non-centrality Parameter (NCP)</td>
<td>0.0</td>
</tr>
<tr>
<td>90 Percent Confidence Interval for NCP</td>
<td>(0.0 ; 11.93)</td>
</tr>
</tbody>
</table>

- Degrees of Freedom = 15
- Minimum Fit Function Chi-Square = 14.20 (P = 0.51)
- Normal Theory Weighted Least Squares Chi-Square = 14.07 (P = 0.52)
- Estimated Non-centrality Parameter (NCP) = 0.0
- 90 Percent Confidence Interval for NCP = (0.0 ; 11.93)

For the second test, we specify a model with equality constraints on the factor loadings.

For the second test, we specify a model with equality constraints on the factor loadings.
Model with Metric Invariance

- Are Factor Loadings Consistent Over Time?
- Test Nested Model
  - Nested Model is Identical to the Baseline Model Except More Constraints are Added
  - Specify Equality Constraints for Factor Loadings
    - Same Factor Loading (a) for BILOOK1, BILOOK2, & BILOOK3
    - Same Factor Loading (b) for BIEIXIT1, BIEIXIT2, & BIEIXIT3

For this second model, we constrain elements in the LAMBDA-Y (LY) matrix to be the same across the three occasion factors.

**EQUALITY CONSTRAINTS ON FACTOR LOADINGS**

<table>
<thead>
<tr>
<th>WC TIME 1 (COLUMN 1)</th>
<th>WC TIME 2 (COLUMN 2)</th>
<th>WC TIME 3 (COLUMN 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>THINKQ1 (ROW 1)</td>
<td>λ₁₁ = 1</td>
<td>0</td>
</tr>
<tr>
<td>BILOOK 1 (ROW 2)</td>
<td>λ₂₁ = 0</td>
<td>0</td>
</tr>
<tr>
<td>BIEIXIT1 (ROW 3)</td>
<td>λ₃₁ = 0</td>
<td>0</td>
</tr>
<tr>
<td>THINKQ2 (ROW 4)</td>
<td>0</td>
<td>λ₄₂ = 1</td>
</tr>
<tr>
<td>BILOOK 2 (ROW 5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BIEIXIT2 (ROW 6)</td>
<td>0</td>
<td>λ₅₂ = 0</td>
</tr>
<tr>
<td>THINKQ3 (ROW 7)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BILOOK 3 (ROW 8)</td>
<td>0</td>
<td>λ₆₂ = 0</td>
</tr>
<tr>
<td>BIEIXIT3 (ROW 9)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>λ₇₃ = 1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>λ₈₃</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>λ₉₃</td>
</tr>
</tbody>
</table>
Thus, we write the following LISREL commands below using the EQ command to constrain factor loadings.

```
TESTING LONGITUDINAL MEASUREMENT INVARIANCE FOR MEASURES OF WITHDRAWAL COGNITIONS
asu working students panel survey - 3 waves
PROGRAM: WITCOG1B.LS8
SECOND TEST OF METRIC INVARIANCE
DA NI = 9 NO = 290 MA = CM
RA = C:WITHCOG.PSF
MO NY = 9 NE = 3 LY = FR TE = SY, FR PS = SY, FR TY = FR AL = FR
LE
WCGTIME1 WCGTIME2 WCGTIME3
PA LY
1 0 0
1 0 0
0 1 0
0 1 0
0 1 0
0 0 1
0 0 1
0 0 1
FI LY 11 LY 42 LY 73
VA 1 LY 11 LY 42 LY 73
EQ LY 21 LY 52 LY 83
EQ LY 31 LY 62 LY 93
FI TY 1
VA 0.0 TY 1
FI TY 4
VA 0.0 TY 4
FI TY 7
VA 0.0 TY 7
PA TE
*
1
0 1
0 0 1
1 0 0 1
0 1 0 0 1
0 0 1 0 0 1
1 0 0 1 0 0 1
0 1 0 0 1 0 0 1
0 0 1 0 0 1 0 0 1
OU SS SC SE TV IT = 600
```

The following reports output for the metric invariance test. Factor loadings in the second model are constrained to be equal across occasions.
The next table reports the overall fit of Model 2. Because this model is nested within Model 1, we can use the $\chi^2$ difference test to gauge model differences as well as the difference in CFI to gauge “practical” differences in fit (Cheung & Rensvold, 2002). Model 2 fits data well and is not different from Model 1. Nonexistent model differences constitute evidence for metric invariance.

<table>
<thead>
<tr>
<th>Models</th>
<th>Model Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>1. Configural invariance</td>
<td>14.20</td>
</tr>
<tr>
<td>2. Metric invariance</td>
<td>21.03</td>
</tr>
<tr>
<td>Difference between Models 1 &amp; 2</td>
<td>6.83</td>
</tr>
<tr>
<td>3. Scalar invariance</td>
<td>53.63*</td>
</tr>
<tr>
<td>Difference between Models 2 &amp; 3</td>
<td>32.60*</td>
</tr>
</tbody>
</table>

*p < .05
To test scalar invariance, we specify a third model that requires intercepts to be constant over time. For this model, we impose equality constraints on entries in the TAU-Y (TY) vector, using the EQ command.

**TESTING SCALAR INVARIANCE**

**ARE INTERCEPTS STABLE OVER TIME?**

- **EQUATE INTERCEPTS FOR THE SAME INDICATORS**
  
- **SAME INTERCEPT FOR BILOOK INDICATORS**
  \[ T_2 = T_5 = T_8 \]
- **SAME INTERCEPT FOR BIEXIT INDICATORS**
  \[ T_3 = T_6 = T_9 \]

Finally, we run the following LISREL program.

```plaintext
TESTING LONGITUDINAL MEASUREMENT INVARIANCE FOR MEASURES OF WITHDRAWAL COGNITIONS
asu working students panel survey - 3 waves
PROGRAM: WITCOG2B.LS8
THIRD TEST OF SCALAR INVARIANCE
DA NI = 9 NO = 290 MA = CM
RA = C:\WITHCOG.PSF
MO NV = 9 NE = 3 LY = FR TE = SY, FR PS = SY, FR TY = FR AL = FR
LE WCGTIME1 WCGTIME2 WCGTIME3
PA LY
1 0 0
1 0 0
1 0 0
0 1 0
0 1 0
0 1 0
0 0 1
```
Not surprisingly, the LISREL output shows identical intercepts for the same indicators.

\[
\begin{array}{ccccccc}
\text{TAU-}\text{Y} & & & & & & \\
\text{THINKQ1} & \text{BILOOK1} & \text{BIEXIT1} & \text{THINKQ2} & \text{BILOOK2} & \text{BIEXIT2} \\
- & - & 0.02 & -0.40 & 0.02 & -0.40 \\
 & & (0.15) & (0.17) & (0.15) & (0.17) \\
 & & 0.13 & -2.33 & 0.13 & -2.33 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
\text{TAU-}\text{Y} & & & & & & \\
\text{THINKQ3} & \text{BILOOK3} & \text{BIEXIT3} & \\
- & - & 0.02 & -0.40 \\
 & & (0.15) & (0.17) \\
 & & 0.13 & -2.33 \\
\end{array}
\]
Though this model significantly differs from Model 2 (significant $\chi^2$ test), it is not materially different according to the CFI difference (.01). Based on Cheung and Rensvold’s (2002) simulation tests, nested models must have CFIs that differ by more than .01 to be truly different. This last test suggests that these measures also have scalar invariance. In summary, our example shows that these measures of withdrawal cognitions have met the necessary preconditions for measurement equivalence. Given these findings, we can next assess change with these measures using SOF LGM.

Though Vandenberg and Lance (2000) identified other psychometric properties that can be tested in invariance tests (e.g., equal uniquenesses), LGM applications mostly check configural and metric invariance to test for time-invariance (Bentein, Vandenberg, Vandenberghe, & Stinglhamber, 2005; Chan & Schmitt, 2000; Garst, Friese, & Molenaar, 2000; Lance et al., 2000). Indeed, Chan (1998) rejects testing uniqueness invariance ($\theta_e$) as being too stringent because changing true variances may underlie “unstable” variances. Nonetheless, scholars interested in tracking mean changes in a construct over time should also test for scalar invariance before assessing mean changes.

**SOF LGM Modeling**

SOF LGM resembles second-order CFA because it specifies growth factors as second-order factors, which in turn, “influence” first-order occasion factors. A fixed set of second-order factor loadings (relating first to second-order factors) then define the growth factors. All occasion factors have a fixed loading of 1 on the intercept factor and have fixed loadings of numbers in an ascending order on the slope factor (defining a linear trajectory). At the same time, each occasion factor “determines” a set of indicators in a given time period. Unlike multilevel analyses, SOF LGM can control for random measurement error when estimating dynamic relationships. Though our illustration uncovered full measurement invariance, SOF LGM can model – and also control for – partial measurement invariance. These are
impressive advantages if researchers suspect that beta or gamma change may bias multiwave data.

Second-Order Factor LGM Model

Outcome

SOF LGM Illustration

Once again, we use our running example with working students. We assessed their withdrawal cognitions on three occasions (Surveys 1–3). We also measured their experiences of shocks – disagreeable events at work that stimulate thoughts of leaving (Mitchell & Lee, 2001) with Survey 1 and turnover (a fourth survey collected after Survey 3). Thus, we test the following SOF LGM model, which posits that shocks increase the initial status of withdrawal cognitions and the trajectory of change for these cognitions. The initial status and slope for withdrawal cognitions in turn should increase the likelihood of quitting.
The three indicators of withdrawal cognitions (e.g., THINKQ1, BILOOK1, BIEXIT1) from each wave define an occasion factor (e.g., WCGTIME1) for that time period. We also estimate autocorrelated errors, allowing for errors for a particular indicator from different periods to co-vary. The first-order factors in turn define the growth factors, using the prescribed fixed factor loadings to define an intercept and slope for withdrawal cognitions. Measured by two indicators, shocks (treated as an exogenous variable in this model), is posited to have direct effects on withdrawal cognitions’ slope and intercept. Finally, the growth factors will have direct effects on turnover (assessed with one indicator).

We use the following data format for this SOF LGM test. In typical LISREL format, indicators of endogenous (occasion, growth, and turnover) factors are positioned in the leftmost columns of a data file. Indicators of the exogenous factor (shocks) are recorded in the rightmost columns.
Our LISREL command statements specify the number of endogenous indicators (10; nine measures of withdrawal cognitions and one quit index), endogenous latent factors (6; three occasion factors, two growth factors, and turnover), exogenous indicators (2; two shock measures), and exogenous latent factor (one shock factor).

We first specify the measurement model for the exogenous factor. Thus, we declare the LAMBDA-X (LX) matrix as full (FU) and fixed (FI).
(the latter sets all loadings to zero). Next, we choose one indicator as a reference indicator (*WORDS*) by fixing its factor loading to 1 and its intercept to 0. We free one LX element so that we can estimate the factor loading for the second SHOCK indicator. Finally, we specify uncorrelated errors for the SHOCK measures.

Next we specify the mean and variance of the exogenous factor and how it influences endogenous factors. Specifically, we free certain parameters in the GAMMA (GA) matrix to estimate how shocks impact withdrawal cognitions’ slope and intercept. We also estimate the variance of the exogenous factor by freely estimating the sole PHI (PH) element. Further, we estimate its mean by freely estimating the sole KAPPA (KA) element.
Then, we specify the measurement model for the endogenous factors. For the longitudinal measurement model, we specified three columns for the LAMBDA-Y (LY) matrix to represent the three occasion factors for withdrawal cognition indicators. For this SOF LGM model, we add three more columns for two growth factors and turnover (and an extra row for the QUIT indicator). The PA command specifies which LY elements are initially fixed (at zero) and which are freed. We next fix factor loadings for reference indicators (namely, THINKQ1, THINKQ2, THINKQ3, and QUIT) by setting their factor loadings to 1. Because previous measurement invariance tests certified metric invariance, we also constrained the freed factor loadings to be the same across occasions using the EQ command (e.g., impose equality constraints on BIEXIT1, BIEXIT2, and BIEXIT3 loadings).
We further specify that reference indicators for endogenous factors have zero intercepts by fixing their entries in the TAU (TY) matrix to be zero. Because earlier invariance tests established scalar invariance, we also constrain the freed intercepts to be equal over time.

**SPECIFY TY MATRIX**

<table>
<thead>
<tr>
<th>MO...TY=FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI TY 10</td>
</tr>
<tr>
<td>VA 0 TY 10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FI TY 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA 0 TY 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FI TY 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA 0 TY 4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FI TY 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA 0 TY 7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EQ TY 3 TY 6 TY 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ TY 2 TY 5 TY 8</td>
</tr>
</tbody>
</table>

---

**SPECIFY MEASUREMENT MODEL FOR ENDOGENOUS FACTORS**

<table>
<thead>
<tr>
<th>LE WC GTIME1 WC GTIME2 WC GTIME3 ALPH A BETA TURN OVER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA LY 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VA 1 LY 10 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI LY 11 LY 4 2 LY 73</td>
</tr>
<tr>
<td>VA 1 LY 11 LY 4 2 LY 73</td>
</tr>
</tbody>
</table>

**REFERENCE INDICATOR FOR TURNOVER (QUIT)**

**SET REFERENCE INDICATORS FOR OCCASION FACTORS**

**METRIC INVARIANCE**

<table>
<thead>
<tr>
<th>VAR1 VAR2 VAR3 VAR4 VAR5 VAR6 VAR7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

| EQ LY 2 1 LY 5 2 LY 8 3 |
| EQ LY 3 1 LY 6 2 LY 9 3 |

---

$\begin{align*}
\lambda_{11} = 1 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{21} = 1 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{31} = 0 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{42} = 1 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{52} = 0 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{62} = 0 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{73} = 1 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{83} = 0 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{93} = 0 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\lambda_{10,6} = 1 & & 0 & & 0 & & 0 & & 0 & & 0 \\
\end{align*}$
Next, we specify elements of the PSI (PS) matrix to estimate variances and covariances among endogenous factors. For identification purposes, we constrain the occasion factors to have zero covariances with each other and all other endogenous variables, though we estimate their variances (e.g., $\psi_{11}$). Moreover, we estimate the intercept ($\psi_{44}$) and slope ($\psi_{55}$) variances as well as their covariance ($\psi_{45}$). Finally, we estimate the “residual” variance for turnover ($\psi_{66}$) – or the variance of its disturbance term (the amount of turnover variance not explained by growth factors).

Then we specify elements of the BETA (BE) matrix to estimate structural relationships among endogenous factors as well as to define the growth factors. To define the intercept, we specify that the intercept factor (BE column 4) have the same structural effect (1) on the three occasion factors (BE rows 1–3). To define the slope, we specify that the slope factor (BE column 5) have a structural effect of 0 on the first occasion factor (WCGTIME1), a structural effect of 1 on the second occasion factor, and an effect of 2 on the third occasion factor. Finally, we free two BE elements to assess structural effects of the growth factors on turnover.
Further, we estimate means for the growth factors and turnover by freeing certain elements in the ALPHA (AL) vector.

**SPECIFY ALPHA VECTOR**

MO.....AL=FI

FR AL 4 AL 5 AL 6

**ALPHA**

\[ \begin{array}{cccccc}
\alpha_1 &=& 0 & \alpha_2 &=& 0 & \alpha_3 &=& 0 & \alpha_4 &=& \alpha_5 &=& \alpha_6 \\
\end{array} \]

Finally, we specify the THETA-EPILSON (TE) matrix to estimate the error variances for endogenous indicators and error autocorrelations. Because turnover is assessed with only one indicator, we fix the error variance of its indicator \( \textit{(QUIT)} \) to zero. In other words, we assume...
that turnover is measured with perfect reliability given only one indicator.

**SPECIFY TE MATRIX**

```
MO.....TE=SY,FR
PA TE *
  1
  0 1
  0 0 1
  1 0 0 1
  0 1 0 0 1
  0 0 1 0 0 1
  1 0 0 1 0 1
  0 1 0 0 1 0 1
  0 0 1 0 0 1 0 1
  0 0 0 0 0 0 0 0 1
FI TE 10 10
VA 0 TE 10 10
```

ESTIMATE MEASUREMENT ERROR VARIANCES AND ERROR COVARIANCES (AUTOCORRELATED ERRORS)

SPECIFY THAT MEASUREMENT ERROR VARIANCE FOR QUIT (Y10, THE SOLE INDEX OF THE TURNOVER CONSTRUCT) = 0. BECAUSE THERE IS ONLY ONE INDICATOR FOR THIS FACTOR, ASSUME THAT THE QUIT INDICATOR HAS PERFECT RELIABILITY (AND THUS NO MEASUREMENT ERROR VARIANCE).

The following LISREL command statement summarizes these specifications of the LISREL matrices and are used for SOF LGM.

```
LINEAR TRAJECTORY MODEL FOR WITHDRAWAL COGNITIONS
DA NI =12 NO =290 MA = CM
RA='C:\WITHCOG2.PSF'
MO NY=10 NE=6 NX=2 NK=1 LX=FU,FI TD=DI,
FR PH=SY,FR GA=FU,FI TX=FR KA=FR LY=FR
TE=SY,FR PS=SY,FI TY=FR AL=FI BE=FU,FI
LK
SHOCKS
VA 1 LX 1 1
FR LX 2 1
FI TX 1
VA 0 TX 1
FR GA 4 1 GA 5 1
LE
WCGTIME1 WCGTIME2 WCGTIME3 ALPHA BETA TURNOVER
PA LY
1 0 0 0 0 0
1 0 0 0 0 0
```
This program yields the following input. Results for the factor matrices show that all factor loadings are significant and that loadings for withdrawal cognition indicators are constant over time (by specification).
Of greater interest are results for the BE estimates, which show that the withdrawal cognitions intercept significantly increased quitting. (We note that turnover is a binary variable, which violates statistical assumptions of interval scaling behind the maximum likelihood estimate procedure used to estimate the model. Another estimation procedure can be deployed in LISREL to correct for this dichotomous criterion.)
It is also instructive to review GA results, which show how the exogenous variable impacts the growth factors. This output reports that shocks increased the initial withdrawal cognitions; students who experienced shocks on their part-time job expressed greater desire to leave this job. Shocks did not, however, influence the rate of change in withdrawal cognitions.

**GA OUTPUT**

<table>
<thead>
<tr>
<th>GAMMA</th>
<th>SHOCKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCGTIME1</td>
<td>- -</td>
</tr>
<tr>
<td>WCGTIME2</td>
<td>- -</td>
</tr>
<tr>
<td>WCGTIME3</td>
<td>- -</td>
</tr>
<tr>
<td>ALPHA</td>
<td>0.03</td>
</tr>
<tr>
<td>(0.00)</td>
<td>6.06</td>
</tr>
<tr>
<td>BETA</td>
<td>0.00</td>
</tr>
<tr>
<td>(0.00)</td>
<td>-1.41</td>
</tr>
<tr>
<td>TURNOVER</td>
<td>- -</td>
</tr>
</tbody>
</table>

Finally, we note the results for the PSI (PS) matrix. We observe that there is significant individual variation in growth factors: the initial status and rate of change for withdrawal cognitions vary across individuals. Growth factors are not related. Further, there remains significant residual variation in turnover that is not explained by the initial status and slope for withdrawal cognitions.
### PS OUTPUT

<table>
<thead>
<tr>
<th>PSI</th>
<th>WCGTIME1</th>
<th>WCGTIME2</th>
<th>WCGTIME3</th>
<th>ALPHA</th>
<th>BETA</th>
<th>TURNOVER</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCGTIME1</td>
<td>0.14</td>
<td></td>
<td></td>
<td>0.07</td>
<td>-</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td></td>
<td>(0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.51</td>
<td></td>
<td></td>
<td>3.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WCGTIME2</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-4</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.31</td>
<td></td>
<td>2.06</td>
</tr>
<tr>
<td>WCGTIME3</td>
<td></td>
<td></td>
<td>-</td>
<td>0.01</td>
<td>-</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.16</td>
<td></td>
<td>11.86</td>
</tr>
<tr>
<td>ALPHA</td>
<td></td>
<td></td>
<td></td>
<td>0.36</td>
<td>-</td>
<td>6.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.06)</td>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BETA</td>
<td></td>
<td></td>
<td></td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.31</td>
<td></td>
<td>2.06</td>
</tr>
<tr>
<td>TURNOVER</td>
<td></td>
<td></td>
<td></td>
<td>0.16</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.86</td>
<td></td>
<td>11.86</td>
</tr>
</tbody>
</table>

According to various fit statistics, this SOF LGM model also fits data well: NFI = .99; CFI = 100; IFI = 1.00; and standardized root mean square residual = .037.

We summarize our suggestions in the following chart, borrowing suggestions from Vandenberg and Lance (2000) and others (Chan, 1998, 2002).
Finally, we note that LGM can accommodate different forms of change, though we illustrated only a linear trajectory of change (Lance et al., 2000) and that this method does not require equal spacing of assessments (Hom et al., 2006).

Comparison of Two Approaches for Assessing Change

Multilevel analyses (e.g., HLM) and latent growth modeling yield similar conclusions (Bliese & Ployhart, 2002), but they have different strengths and weaknesses. A key strength of multilevel approaches is their more flexible data requirements; they can handle multiwave data with differing number of measures and timing of measurement for entities. By contrast, LGM requires equal number and spacing of assessments for all entities: balanced and time-structured data (Duncan et al., 1999; Singer & Willett, 2003). LGM can, however, more readily investigate dynamic relationships between variables, such as relationships between two variables’ change trajectory (cf. Chan & Schmitt, 2000). Requiring more measures per wave, SOF LGM can control for random errors of measurement as well as partial measurement invariance. These controls would improve estimates of mean changes and dynamic relationships. Armed with this procedural knowledge, organizational researchers can revisit dynamic phenomena. Exploring temporal changes not only can provide stronger evidence for causality (Dwyer, 1983) but also establish that the rate of change of a predictor variable can have additional explanatory power beyond its static value (Harrison, Virick & William, 1996).

REFERENCES


Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). Organizational Research Methods, 1, 421–483.


INTERPRETING EMPIRICAL RESULTS IN STRATEGY AND MANAGEMENT RESEARCH

J. Myles Shaver

ABSTRACT

As a field, we should put more emphasis on interpreting the magnitude of coefficient estimates rather than only assessing statistical significance. To support this claim, I demonstrate how focusing only on statistical significance can lead to incorrect and incomplete conclusions in many common applications of the linear regression model. Moreover, I demonstrate why interpreting coefficient estimates in common non-linear estimators (e.g., probit, logit, Poisson, and negative binomial estimators) requires additional care compared to the linear regression model.

A great deal of empirical research in strategy and management revolves around statistical inference. That is, we want to make statements about a population based on descriptions of a sample. Most econometric textbooks highlight two related elements of inference in classical statistics: estimation and hypothesis testing (e.g., Kmenta, 1986; I appreciate helpful comments from Wilbur Chung, Gary Dushnitsky, Adam Fremeth, Mazhar Islam, Arturs Kalnins, Rob Salomon, David Souder, and Minyuan Zhao.)
Greene, 2000). With estimation, we go to the data to seek descriptions of the sample. That is, we estimate parameters to describe the sample. With hypothesis testing, we assess if our prior belief about the population (i.e., our maintained hypothesis) is consistent with the estimated parameters. Naturally, estimation and hypothesis testing are intertwined. We use properties of the estimators that we employ to test if our maintained hypothesis is refuted.

As a field, we are overly focused on the hypothesis testing aspect of inference and this comes with the cost of missed opportunity to better understand the phenomena that we work so hard to study. Many discussions of empirical findings in our field either exclusively or overly rely on reporting statistical significance in the discussion of the results. At the extreme, one could delete all but the asterisks (*) in the results tables and the presented interpretation would not suffer.

My goal in this chapter is to argue why we should further discuss the magnitude of the parameters we estimate (i.e., effect size) and how to approach this task. I do not assert that we should forgo tests of hypotheses. Rather, my contention is that we miss an opportunity to more fully understand the phenomena that we study when we focus exclusively on reporting hypothesis tests.

The structure of the chapter is as follows. I begin with a discussion of the linear regression model because this is the basis for a large volume of statistical inference in strategy and management. Within the linear regression model, I show exactly what coefficient estimates represent when it comes to interpreting empirical findings. This will be consistent with the reader’s intuition. I then highlight several situations where focusing on statistical significance and ignoring the magnitude of coefficient estimates either: (a) misses opportunities to understand the nature of the relationships we estimate or (b) reaches incorrect conclusions about the nature of the relationship that we estimate.

From here, I highlight that coefficient estimates do not share the same interpretation in non-linear regression models (e.g., logit, probit, Poisson, and negative binomial regression models) that they do in linear models. Therefore, interpreting empirical findings from these models requires additional care.

In making the following points, I present some very simple calculus. Many of the points I wish to make become clearer when I can demonstrate their origin and presenting the equations facilitates this. I realize that this comes with the potential cost of insulting the mathematically apt reader and alienating the non-mathematically inclined reader – neither of which is my goal. For the mathematically inclined reader, I hope the interpretations and examples add value and the equations make the structure transparent. For the
non-mathematically inclined reader, I believe that the text presents the arguments and the examples will illustrate the points. You can refer to the equations if you desire to see what motivates and structures the conclusions that I draw.

WHAT DO REGRESSION COEFFICIENT ESTIMATES REPRESENT IN THE LINEAR REGRESSION MODEL?

A large proportion of empirical analysis in strategy and management research utilizes the linear regression model. Standard practice when reporting results from these models is to present coefficient estimates and assess if they are statistically significant (i.e., perform the hypothesis test that they differ from zero). Independent variables with coefficients that are statistically significant are said to affect the dependent variable. Thus, the commonly held intuition is that coefficient estimates capture how independent variables affect the dependent variable.

Such intuition is accurate. To be more formal, regression coefficients in the linear regression model are partial derivatives or marginal effects of the predicted values from the estimation. That is, coefficient estimates represent the change in the expected value of the dependent variable that corresponds to a change in the independent variable – holding all other factors constant. To see this, consider a standard regression model as depicted in Eq. (1).

\[ Y = a + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \] (1)

Once we obtain estimates from Eq. (1), we can assess how the independent variables affect the expected value of the dependent variable (i.e., the predicted value). Eq. (2) presents the expected value of Eq. (1).

\[ E[Y] = a + \beta_1 x_1 + \beta_2 x_2 \] (2)

Eqs. (3) and (4) evaluate the partial derivative of the expected value of \( Y \) with respect to \( x_1 \) and \( x_2 \), respectively.

\[ \frac{\partial E[Y]}{\partial x_1} = \beta_1 \] (3)

\[ \frac{\partial E[Y]}{\partial x_2} = \beta_2 \] (4)
As one can see, the coefficient estimate $\beta_1$ represents the effect of changing $x_1$ while holding $x_2$ constant. Therefore, if $x_1$ increases by 1, $Y$ changes by the value of $\beta_1$. Likewise, coefficient estimate of $\beta_2$ represents the effect of changing $x_2$ while holding $x_1$ constant. If $x_2$ increases by 1, $Y$ changes by the value of $\beta_2$.

**Beyond Hypothesis Tests**

Although hypotheses tests are important when assessing our theories, there is also information in the magnitude and nature of the coefficient estimates that can further inform this assessment. I examine a number of situations where reliance on hypothesis tests without assessing the magnitude and nature of the coefficient estimates can result in incomplete or incorrect inference.

**Are Non-Zero Coefficient Estimates Actually Meaningful?**

The standard hypothesis test we report is the test with a null hypothesis that the coefficient estimate equals zero.$^1$ Statistically significant tests reject the null hypothesis and suggest that the coefficient estimate is non-zero. The interpretation is – provided we have satisfied the assumptions underlying the test – the probability that the true value of the coefficient estimate in the population equals zero is less than the cut-off of our test. For example, if the $p$-value of a coefficient estimate is 0.02, then the probability that the true value of coefficient estimate in the population equals zero is 2 percent.

There are two important considerations for such tests. First, is the appropriate null hypothesis that the coefficient estimate equals zero? Second, even if the coefficient estimate is statistically significant, is it of a magnitude that it has “organizational significance”?

Consider the first question. The standard output from statistical software packages presents hypothesis tests where the null hypothesis is zero. However, I have reviewed many papers where the null hypothesis should not have been zero and the authors have relied on the hypothesis test information from the statistical output. For example, if the authors predict why the dependent and independent variables do not move in parallel to each other, then the appropriate null hypothesis is that the coefficient estimate is 1. Therefore, in some situations the null hypothesis should reflect the magnitude in which an independent variable affects a dependent variable.
Nevertheless, in many circumstances the appropriate null hypothesis is that the coefficient estimate equals zero (i.e., the independent variable does not affect the dependent variable). In these cases, if the coefficient estimate is statistically significant, it still begs the question of whether the non-zero effect of a magnitude that is important. It is possible that a coefficient estimate tests different from zero yet has little practical impact on the phenomena of study (for a further discussion of this issue see McCloskey & Ziliak, 1996). However, in our literature little attention is paid to the magnitude of an effect that tests different from zero.

To illustrate this point, I simulate data to highlight the difference between statistical significance and the magnitude of coefficient estimates. The advantage of simulating data is that I determine and hold constant the underlying structure in the data. This makes it possible to identify what causes any observed changes across simulations.

Table 1 presents three sets of regression results from simulated data with the following structure. The average value of the dependent variable ($Y$) is approximately 500, the independent variable ($x$) ranges from 300 to 450, and the marginal effect of this variable on the dependent variable is 0.01. Therefore, moving the independent variable across its entire range changes the dependent variable only 1.5 when its average value is approximately 500.

Is the effect important? I would argue that there are reasons to believe that this effect is not practically important. Moving the dependent variable with mean of 500, 1.5 does not reflect much change and would likely have little practical importance. However, the regression results in Model 1.1 show a statistically significant effect of $x$ on $Y$.

Why would we see a statistically significant result when the magnitude of the effect is so small? With a larger sample we get more powerful tests to conclude if the estimates differ from zero. However, at some point differing from zero in a practical sense might not be meaningful. (For a further analysis of this issue see Leamer, 1978, Chapter 4.)

To see this in the data, revisit Table 1. Notice that the level of statistical significance falls across Models 1.1–1.3 as the sample size decreases, even though the magnitude of the estimate is actually slightly larger in Models 1.2 and 1.3. Here it is clear that statistical significance is being driven by changes in sample size – not changes in the underlying data structure. What we can appropriately conclude from the increase in the level of statistical significance is the probability that the sample in Model 1.1 comes from an underlying population where the true effect of $x$ on $Y$ is 0 is lower than the sample in Model 1.2 or 1.3. The effect is not stronger because the magnitude of the coefficient estimate is actually smaller.
The above discussion highlights why focusing on tests of statistical significance can obscure important information with respect to inference. In particular, the discussion highlights how statistically significant results can reflect effects that are not meaningful and, as such, might not be consistent with underlying predictions.

Does the Value of the Coefficient Estimate Have “Face Validity”? Another reason why it is important to assess the magnitude of coefficient estimates is to aid assessing whether the underlying theoretical mechanisms
drive the estimated effect. In particular, it is possible that coefficient estimates take values that appear “too large” to be driven by an underlying theoretical mechanism. Whereas the previous discussion highlighted how focusing on statistical significance can lead us to accept results that are not meaningful, focusing on statistical significance can also prevent us from questioning results with magnitudes that are much larger than we would expect.

Consider the following hypothetical example. Suppose we examine the stock market reaction of CEO turnover within a given industry. We predict that executive turnover is negatively related to stock price in this industry when turnover is involuntary because of organizational distress associated with terminating its leader. We regress abnormal returns from the announcement of CEO turnover on a dummy variable indicating if the turnover was voluntary or not. We find a negative and significant effect and claim support for our hypothesis.

What if on closer examination we find an estimate of $-0.50$, which would mean that the abnormal return was $-50$ percent on average? This would beg the question of whether organizational distress was so large as to almost halve the value of the firm. If we believe this effect is too large, it might suggest that some other effect is working in parallel to the predicted effect. It might also suggest that something else is happening instead of the predicted effect.

Therefore, when we focus only on hypothesis testing and ignore the magnitude of the effects, we suppress information from our estimation efforts. This information can suggest that effect magnitudes are too large to be driven by some mechanisms. Only if we analyze the magnitude of the effects, we can make assessments about the face validity of our results.

*Have We Correctly Interpreted a Non-Linear Effect?*

It is possible to estimate non-linear effects within the linear regression model by including exponential terms in the regression equation. For example, a common situation is where we want to assess if there exists a U-shaped or inverted U-shaped relationship of $x$ on $Y$. This can be accomplished by adding a squared term of the independent variable into the equation.

Assessing the magnitude of coefficient estimates is critical for making valid inference in these situations because it provides information whether the curve is U-shaped or inverted U-shaped, and identifies where the curve inflects. Depending where the point of inflection is relative to the data in the sample, a curve that increases or decreases over the entire range of the
sample can mistakenly be interpreted as a U-shaped or inverted U-shaped curve.

To see this, let us start with the standard set-up for specifying the test for a U-shaped or inverted U-shaped relationship in OLS. Here, we regress the dependent variable \( Y \) on the independent variable \( x \) and its square \( x^2 \). The expected value of this equation is represented by Eq. (5).

\[
E[Y] = \alpha + \beta_1 x + \beta_2 x^2
\]

(5)

The partial derivative of \( Y \) with respect to \( x \) (i.e., the marginal effect of \( x \) on \( Y \) ) is the following:

\[
\frac{\partial E[Y]}{\partial x} = \beta_1 + 2\beta_2 x
\]

(6)

From Eq. (6), we can see that the marginal effect of \( x \) on \( Y \) depends on the value of \( x \), in addition to the coefficient estimates \( \beta_1 \) and \( \beta_2 \). This makes sense because in order to generate a U-shaped curve, the marginal effect on \( Y \) of increasing \( x \) initially has to be negative, increase in value (i.e., become less negative) become positive, and then become increasingly positive as \( x \) increases. Because Eq. (5) can test the existence of either U-shaped or inverted U-shaped curves, the curvature of the curve is evaluated by the second derivative with respect to \( x \).

\[
\frac{\partial^2 E[Y]}{\partial x^2} = 2\beta_2
\]

(7)

If \( \beta_2 > 0 \) then the curve is convex and potentially U-shaped and if \( \beta_2 < 0 \) then the curve is concave and potentially inverted U-shaped. Moreover, the curve inflects at the point where the marginal effect (i.e., Eq. (6)) equals zero. Therefore, the inflection point is where \( x \) takes the following value.

\[
x = \frac{-\beta_1}{2\beta_2}
\]

(8)

Testing for a U-shaped or inverted U-shaped curve in this manner imposes a quadratic specification (i.e., we include \( x \) and \( x^2 \) in the specification). Therefore, the values of the parameter estimates reflect the shape of the quadratic function that best fits the data. It does not, however, prove that this is the underlying form of the data. To better appreciate this point, consider the data presented in Fig. 1.

The data in Fig. 1 come from an underlying function where \( Y \) increases as \( x \) increases – albeit at a decreasing rate. It is a non-linear function – but it is
not inverted U-shaped because \( Y \) does not ever decrease as \( x \) takes larger values. If I were to approach these data and estimate a quadratic specification as is standard practice, I would get the empirical estimates presented in Model 2.1 of Table 2.

Commonly presented arguments for the existence of an inverted U-shaped relationship in our literature are when a model such as 2.1 provides the following estimates: (a) \( \beta_2 \) is negative and significant and (b) \( \beta_1 \) is positive and significant. This is not an accurate justification for the following reasons. First, negative values of \( \beta_2 \) only indicate that the estimated curve is concave. If the curve does not inflect over the range of the data, there will not be an inverted U. Therefore, a negative value of \( \beta_2 \) is a necessary but not sufficient condition for an inverted U-shaped curve. Second, if \( \beta_2 \) is negative, then a positive estimate of \( \beta_1 \) only indicates that the curve inflects where \( x \) takes a positive value. However, if the range of \( x \) includes negative values, an inverted U-shaped curve could inflect where \( x \) takes a negative value and there would be no reason to expect that \( \beta_1 \) be positive.

Returning to the estimates in Model 2.1, I find a negative and significant estimate of \( \beta_2 \) and a positive and significant estimate of \( \beta_1 \). However, the magnitude of the coefficient estimates provides evidence that the estimated curve is not inverted U-shaped. The negative estimate of \( \beta_2 \) indicates that the curve is concave. The inflection point as highlighted by Eq. (8) is

---

**Fig. 1.** Graph of Data and Fitted Curve from Simulation 2 – Model 2.1.
Table 2. Simulation 2
Simulated data reflect the underlying functional form:
\[ Y = 100 + 50 \ln(x) + \varepsilon \]
Estimated specification in Model 2.1: \[ Y = \alpha_1 + \beta_1 x + \beta_2 x^2 + \varepsilon_1 \]
Estimated specification in Model 2.2: \[ Y = \alpha_2 + \beta_3 \ln(x) + \varepsilon_2 \]
(\( t \)-values in parentheses).

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>Model 2.1</th>
<th>Model 2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>3.507***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(13.03)</td>
<td></td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.027***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-7.38)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>188.99***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(42.86)</td>
<td></td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td></td>
<td>49.98***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(34.64)</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td></td>
<td>99.97***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(19.73)</td>
</tr>
</tbody>
</table>

Approximate inflection point: 65
\( R^2 \): 0.82 0.83

Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 274.02 )</td>
<td>24.74</td>
<td>213.46</td>
<td>320.86</td>
<td></td>
</tr>
<tr>
<td>( x = 35.67 )</td>
<td>14.12</td>
<td>11.10</td>
<td>59.85</td>
<td></td>
</tr>
</tbody>
</table>

Note: Stata code for this simulation is presented in the appendix.

\(* p < 0.1 \) (two-tailed tests).
\(** p < 0.05 \).
\(*** p < 0.01 \).

approximately where \( x \) takes the value 65. This point is outside the range of \( x \) in the data (\( x \) ranges from approximately 11 to 59). Therefore, the estimates in Model 2.1 suggest a curve that increases at a decreasing rate, but not an inverted U-shaped relationship.

I should note that it is possible to find inflection points within the data range even when the underlying functional form is not U-shaped or inverted U-shaped.\(^5\) The reason is that imposing a quadratic specification like Eq. (5) generates estimates of the best-fitting quadratic function to the data – not the best-fitting function to the data. Therefore, while finding inflection points outside the data range will indicate that a U-shaped or inverted
U-shaped curve does not exist within the data (as specified); an inflection point within the data range does not prove that a U-shaped or inverted U-shaped curve exists within the data.

**Have We Correctly Interpreted an Interaction Effect?**

Another common test in the linear model that requires careful interpretation of coefficient estimates is the test of interaction effects (i.e., moderating effects). Once again, the standard practice in the strategy and management literature is to interpret statistical significance as support for hypotheses that predict interactions or moderating effects. Although this can often lead to correct conclusions, it can also be misleading.

To illustrate this point let me start with the standard model specification used to estimate interactions effects.

\[
Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \tag{9}
\]

Assume that the theory being tested is that \(x_1\) positively affects \(Y\) with \(x_2\) mitigating this effect. Namely, the positive effect of \(x_1\) on \(Y\) decreases as \(x_2\) increases. Once again, to understand the effect of \(x_1\) on \(Y\) we take the partial derivative of \(x_1\) with respect to the expected value of \(Y\). Eq. (10) highlights this.

\[
\frac{\partial E[Y]}{\partial x_1} = \beta_1 + \beta_3 x_2
\tag{10}
\]

If we want to assess how the effect of \(x_1\) on \(Y\) changes as \(x_2\) changes, we take the partial derivative of Eq. (10) with respect to \(x_2\). This results in the following equation.

\[
\frac{\partial^2 E[Y]}{\partial x_1 \partial x_2} = \beta_3 \tag{11}
\]

Therefore, the test of whether \(\beta_3\) differs from zero is the appropriate test to assess if changing \(x_2\) affects the marginal effect of \(x_1\) on \(Y\). If \(\beta_3\) is non-zero, then the effect of \(x_1\) on \(Y\) depends on the value of \(x_2\). If \(\beta_3\) is zero, then the effect of \(x_1\) on \(Y\) does not depend on the value of \(x_2\). However, testing if \(\beta_3\) differs from zero does not directly inform the magnitude of the effect of \(x_1\) on \(Y\). To assess this, we need to assess Eq. (10).

Suppose that we estimate Eq. (9) and find a positive and significant estimate of \(\beta_1\) and a negative and significant estimate of \(\beta_3\). Most studies would claim support for the underlying theory that predicted \(x_1\) positively affects \(Y\) with \(x_2\) mitigating this effect. However, let us assume that \(\beta_1\) takes the value of 1, \(\beta_3\) takes the value of \(-5\), and \(x_2\) ranges from 2 to 10. From
Eq. (10) we can determine that the marginal effect of $x_1$ on $Y$ ranges from $-9$ (when $x_2 = 2$) to $-49$ (when $x_2 = 10$).

These estimates would not support the claim that $x_1$ positively affects $Y$ with $x_2$ mitigating this effect. Here, $x_1$ would negatively affect $Y$ and this negative effect would strengthen as $x_2$ increases. One might ask, “Wasn’t $\beta_1$ positive and significant?” Yes. However, referring back to Eq. (10), one can see that $\beta_1$ is the marginal effect of $x_1$ on $Y$ only when $x_2$ equals zero – provided that $\beta_3$ is different than zero. Therefore, in order to interpret the nature of an interaction effect one needs to assess the magnitude of the coefficient estimates and the range of the independent variables.

*Have We Correctly Interpreted the Effects of Scaled Variables?*

Another issue that can complicate the interpretation of results and tests of statistical significance is scaling independent variables. This is because the coefficient estimates from these models represent the marginal effect of changing the independent variable – given how it is scaled.

Take, for example, the common situation where an independent variable is scaled by taking its natural logarithm. The coefficient estimate of this variable in the regression equation indicates how the dependent variable changes as the logarithm of the independent variable changes. It does not represent the marginal effect of changing the variable that underlies the scaled independent variable in the regression model (i.e., the variable that was transformed). To see this, consider the estimates presented in Model 2.2 of Table 2, which has the logarithmic transformation of $x$. The expected value of the dependent variable is represented by the following equation.

$$E[Y] = 100 + 50 \ln(x)$$  \hspace{1cm} (12)

The marginal effect of $x$ on $Y$ is found by taking the partial derivative of Eq. (12) with respect to $x$.

$$\frac{\partial E[Y]}{\partial x} = \frac{50}{x}$$  \hspace{1cm} (13)

If we assume that $x$ measures firm size, seeing Model 2.2 one might be inclined to conclude that size significantly affects the dependent variable. However, the more precise conclusion is that the logarithm of size affects the dependent variable. The effect of size on the dependent variable depends on the value of the independent variable. In Eq. (13) increasing $x$ once it is at a large value has practically no effect on $Y$. For example, the marginal effect is 0.01 if $x$ equals 5000.
Scaling a variable by taking its natural logarithm is a relatively straightforward transformation to justify – we expect a concave relationship of $x$ on $Y$. It is also relatively straightforward to interpret the effect of the underlying variable as I demonstrated in the above paragraph. However, not all variable transformations are so straightforward. Moreover, one cannot forget that the coefficient estimate of the transformed value does not identify the marginal effect of that variable on the dependent variable. With such complications in interpreting findings of transformed variables I recommend avoiding this practice unless the transformation is to invoke a functional form where one has a priori an expectation why this form exists. I have often seen transformations of independent variables justified with a statement that the variable of interest is not normally distributed. However, the linear regression model does not require any assumptions about the distribution of independent variables.

Finally, it is worthwhile to discuss one type of scaling that we see in management, although rarely in strategy research. This is standardizing variables in order to estimate standardized coefficient estimates. Standardizing is scaling all variables (dependent and independent) so that they have zero mean and standard deviation of 1. The advantage of standardizing is that it aids greatly in the assessment if the coefficient estimate is of a magnitude that is meaningful. This is because the coefficient estimate now indicates how many standard deviations the dependent variable changes when the independent variable changes by one standard deviation. The disadvantage is that in order to compare results across samples or make inferences outside the sample, one has to assume that the distribution of the variables across samples has to be equal.

**What Descriptive Statistics Should We Provide Readers to Aid Interpreting Our Findings?**

Based on the examples presented above, I hope to have made a convincing case why interpreting coefficient estimates is important to reach informed conclusions. However, to reach these conclusions and to allow readers to assess the interpretation, it becomes necessary to present additional descriptive statistics compared to what is commonly done. The standard of practice in strategy and management research is to present means, standard deviations, and correlations among variables. However, to make many of the assessments in the preceding pages one requires information about the range of the variables and potentially their distribution. Therefore, I would hope that this becomes standard practice – especially, when performing tests of the types described above.
In addition, with the above information it is possible for a reader to graph the relationships in models where the marginal effect changes across the range of the observed data (i.e., U-shaped relationships, inverted U-shaped relationships, and interaction effects). For the non-mathematically inclined reader, graphing such relationships might provide a more intuitive understanding than examining the equations that I have presented.

**INTERPRETING COEFFICIENT ESTIMATES FROM NON-LINEAR ESTIMATORS**

Until now, I have focused discussion on interpreting coefficient estimates in the linear regression model. Non-linear regression models have become increasingly utilized in strategy and management research. Examples of such models include logit, probit, Poisson, and negative binomial regression models. The advantage of these models is that they provide more efficient estimates than linear estimators for certain types of limited dependent variables. However, the increased efficiency comes with the “cost” that more nuanced interpretations of these estimates are required. The following sections illustrate this.

*Coefficient Estimates Do Not Have the Same Interpretation as in the Linear Model*

In the linear regression model, coefficient estimates had the interpretation of partial derivatives (i.e., marginal effects). Referring back to Eq. (3), if we change $x_1$ by 1, $Y$ changes by $\beta_1$. Coefficient estimates do not have the same interpretation in non-linear models. To see this, let me assess the logistic regression model.

A common use of logistic regression is when the dependent variable is dichotomous (0 or 1). The functional form that underlies the model represents the probability that the dependent variable takes the value 1. For simplicity, consider a model with only a constant ($a$) and one independent variable ($x$). Eq. (14) portrays the probability that the dependent variable equals 1 from this logit specification.

$$\text{Prob}(Y = 1) = \frac{e^{a + \beta x}}{1 + e^{a + \beta x}} \quad (14)$$

As with the linear regression model, we are interested in how changes in the independent variable affect the predicted value of the dependent variable.
(i.e., expected value). Because Eq. (14) represents the probability that the dependent variable equals 1, the expected value of the dependent variable is 1 times this probability plus 0 times the probability $Y=0$ (which is $1-\text{Prob}(Y=1)$). Eq. (15) presents this.

$$E[Y] = 1 \left[ \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \right] + 0 \left[ 1 - \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \right] = \left[ \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \right]$$ (15)

Once again, if we want to see how changing $x$ affects the expected value of $Y$, we take the partial derivative of the expected value of $Y$ with respect to $x$ as in Eq. (15).

$$\frac{\partial E[Y]}{\partial x} = \left[ \frac{e^{\alpha + \beta x}}{(1 + e^{\alpha + \beta x})^2} \right] \beta$$ (16)

Notice in Eq. (16) that the coefficient estimate is not the marginal effect as it was with the linear regression model. Namely, we cannot interpret $\beta$ as how the probability that $Y$ equals 1 changes when $x$ increases by 1. Rather, the marginal effect is the coefficient estimate multiplied by the term that precedes it in Eq. (16).6

Intuitively, this can be considered in the following manner. The underlying logit function is non-linear; therefore, the effect of changing $x$ on the expected value of $Y$ will depend where on this non-linear curve the observation lies. The term in Eq. (16) that precedes $\beta$ incorporates this information.

From Eq. (16), one can see why tests, if $\beta$ differs from zero, are important to make statistical inferences. If $\beta$ equals zero, then the marginal effect is zero – regardless the value of the term that proceeds it.

However, if $\beta$ differs from zero it does not provide us with information to assess if the effect of $x$ is meaningful or has face validity. To do this requires calculating the value of Eq. (16). Consider the following example. I estimate a logit model of the above specification. The estimate of $\beta$ is 1 and the estimate of $\alpha$ is 1. When $x=0$, the marginal effect is 0.197 (19.7 percent). When $x=5$, the marginal effect is 0.002 (0.2 percent). In this case, moving along the curve has a large impact on the magnitude of the marginal effect.

Because the marginal effect varies depending on the value of the independent variable (or all independent variables if the specification includes more than one), there are three potential ways to assess the magnitude of the marginal effect. First is to assess the marginal effect at the mean of the independent variables. This would be most appropriate if the mean value provided a meaningful description of the average firm. For
instance, if the independent variables were continuous and normally distributed, then the mean value would reflect the “average” firm. However, if the set of independent variables were all categorical, this would be less appropriate. Second is to calculate the value of Eq. (16) for each observation and then average that across the sample. This provides information of the average effect of changing $x$ on the probability that $Y = 1$ within the sample. Third is to calculate the value of Eq. (16) for particular values of interest. This would be the most appropriate if there were “types” of firms that: (a) the researcher was particular interested in and (b) could be easily represented by sets of values of the independent variables.

In summary, coefficient estimates in non-linear regression models do not have the same interpretation as coefficient estimates in the linear regression model. Moreover, interpreting the magnitude of the effect is more nuanced in non-linear regression models compared to linear models – even for simple relationships. As the following section will highlight the interpreting effects in non-linear models becomes even more nuanced in for relationships like interaction effects.

Interaction Effects become Exceedingly Difficult to Interpret in Non-Linear Models

The discussion of the linear regression model highlighted why interpreting interaction effects had to go beyond looking at the statistical significance of coefficient estimates. This was because interpreting marginal effects required more than a significant coefficient estimate from the interaction term. The following discussion highlights why assessing interaction effects is even more complex with non-linear models.

To illustrate this point, it is useful to go back the discussion surrounding Eqs. (9), (10), and (11). The idea underlying the interaction effect is that the marginal effect of $x_1$ on $Y$ is contingent on the value of $x_2$. To make this assessment, we take the partial derivate of $x_1$ on $Y$ with respect to $x_2$ (i.e., the cross-partial derivative of $Y$ with respect to $x_1$ and $x_2$). In the linear regression model this value was $\beta_3$ (i.e., Eq. (11)). Therefore, testing if the marginal effect of $x_1$ on $Y$ is contingent on the value of $x_2$ is equivalent to testing $\beta_3$.

Unfortunately, the assessment is not so straightforward with non-linear regression models. To illustrate this point, I consider the Poisson regression model. The Poisson regression model provides more efficient estimates than the linear regression model when the dependent variable is count data.
(i.e., integers with a lower bound of zero). My motivation for assessing the Poisson model is to demonstrate the interpretation concerns in another commonly used non-linear model and Ai and Norton (2003) make the parallel assessment for the logit and probit models.

Assume that we want to assess the interaction effect of $x_1$ and $x_2$ on the dependent variable. Therefore, we include $x_1$ and $x_2$ and $x_1x_2$ as independent variables in the model specification. The expected value $Y$ within a Poisson regression model for this specification is represented by the following equation.

$$E[Y] = e^{x_1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2} \quad (17)$$

The marginal effect of $x_1$ on the expected value of $Y$ is assessed by taking the partial derivative of Eq. (17) with respect to $x_1$. This is represented by the following equation.

$$\frac{\partial E[Y]}{\partial x_1} = e^{x_1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2} (\beta_1 + \beta_3 x_2) \quad (18)$$

The difference between Eq. (18) and the parallel assessment in the linear model (Eq. (10)) is that Eq. (10) only includes the bracketed term. The intuition why the term before the bracketed term appears in Eq. (18) is because the Poisson model is a non-linear estimator, the marginal effect will depend where on the non-linear curve an observation lies. This is parallel to the discussion of marginal effects in the logit model.

Continuing with the determination of whether the effect of $x_1$ on $Y$ varies depending on the value of $x_2$, now I take the partial derivative of Eq. (18) (i.e., the marginal effect of $x_1$ on $Y$) with respect to $x_2$, which is represented by Eq. (19).

$$\frac{\partial^2 E[Y]}{\partial x_1 \partial x_2} = e^{x_1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2} (\beta_2 + \beta_3 x_2) (\beta_1 + \beta_3 x_2) \quad (19)$$

The appropriate test for an interaction effect in a Poisson regression model is to test if Eq. (19) differs from zero. This is much different than the linear regression model where the parallel test was that $\beta_3$ differs from zero (Eq. (11)). Moreover, the sign of the coefficient estimate $\beta_3$ does not even have to be consistent with the direction of the interaction. The magnitude and direction of the interaction is a function of all the coefficient estimates and the value of the independent variables.
Assessing an interaction effect in non-linear models, therefore, can be complex. One approach I have found useful to make such comparisons is to split the sample, estimate the effect in each sub-sample, and compare the marginal effects across samples (e.g., Penner-Hahn & Shaver, 2005). This method has the advantage of not having to rely on the above calculation to justify the interaction effect. Although justification of where to split the sample can sometimes be tenuous, in other cases it is straightforward (e.g., when you have a dichotomous variable). Moreover, comparing marginal effects across samples rather than coefficient estimates is important because, as noted above, coefficient estimates in non-linear models do not reflect marginal effects. Therefore, testing differences in coefficient estimates across samples can be misleading.

To summarize, assessing interaction effects is very different in non-linear models compared to linear models. It requires going further than assessing if coefficient estimates are statistically significant. Moreover, if we apply the standard interpretation based on the linear model to estimates from non-linear models then we potentially reach incorrect conclusions.

**CONCLUSION**

My goal in this chapter is to draw attention that hypothesis testing, while central to statistical inference, is not the only consideration when interpreting empirical results. Rather, careful interpretation of empirical results requires that the nature and magnitude of coefficient estimates be considered.

Assessing the magnitude of the effects that we estimate has many desirable outcomes. It can lead to different conclusions and more nuanced interpretations versus relying only on hypothesis tests. Moreover, it also provides a deeper understanding of the underlying phenomena that we study. With the extensive efforts often required to collect data sets in our field, we miss the opportunity to more completely and accurately understand the phenomena we study if we only assess whether independent variables have non-zero effects on the dependent variable.

Fortunately, the solution to this concern is well within grasp and does not require a completely different approach. Rather, it involves: (a) presenting additional descriptive statistics of the variables in the empirical model; (b) relying on hypothesis tests, as is current practice; and (c) taking the time to interpret coefficient estimates, which we currently report in our results’ tables. This additional effort has the potential to pay off many times in
advancing our understanding of the empirical relationships that we currently report.

NOTES

1. This would reflect a two-tailed test. A one-tailed test has the null hypothesis that the coefficient is either (a) equal or greater than zero or (b) equal or less than zero.
2. The appendix provides Stata code used to generate the simulations in this chapter.
3. One will also note that as the sample size increases, the coefficient estimates come closer to the true values of the underlying parameters as expected.
4. The situation that I discuss is distinct, yet related, to the situation where statistical tests lack the power to identify meaningful effects due to small sample sizes (e.g., Ferguson & Ketchen, 1999; Brock, 2003).
5. Interested reader can manipulate the simulation code in the appendix to observe this.
6. Formulas to calculate marginal effects for other non-linear estimators can be found in most econometric texts (e.g., Greene, 2000).

REFERENCES

APPENDIX. Stata Code for Simulated Data

Table A1

notes: set seed of random number generator to ensure reproducibility
set seed 987654321
notes: number of observations
set obs 10000
notes: random number generator N(0,1)
global RNDM invnorm(uniform ( ))
notes: random number generator uniform distribution
global UNI uniform ( )
gen e = $RNDM
gen u = $UNI
gen x = (150 * u) + 300
gen y = 500 + 0.01 * x + 10 * e
sum y x
reg y x
drop if _n > 1000
reg y x
sum y x
drop if _n > 100
reg y x
sum y x

Table A2

notes: set seed of random number generator to ensure reproducibility
set seed 987654321
notes: number of observations
set obs 250
notes: random number generator N(0,1)
global RNDM invnorm(uniform ( ))
notes: random number generator uniform distribution
global UNI uniform ( )
gen e = $RNDM
gen u = $UNI
gen x = 10 + 50 * u
gen y = 100 + 50 * ln(x) + 10 * e
gen x2 = x * x
gen log_x = log(x)
sum y x
reg y x x2
reg y log_x
scatter y x || qfit y x
This page intentionally left blank
MEDIATION IN STRATEGIC MANAGEMENT RESEARCH: CONCEPTUAL BEGINNINGS, CURRENT APPLICATION, AND FUTURE RECOMMENDATIONS

Toyah L. Miller, María del Carmen Triana, Christopher R. Reutzel and S. Trevis Certo

ABSTRACT

Mediating effects allow strategic management researchers to understand “black box” processes underlying complex relationships whereby the effect of an independent variable is transmitted to a dependent variable through a third variable. Since the seminal work of Baron and Kenny (1986), advancements have been made in mediation analysis. Thus, literature on the latest techniques for analyzing mediating and intervening variables is presented. In addition, strategy literature published in the Academy of Management Journal and the Strategic Management Journal between 1986 and 2005 employing tests of mediation is reviewed to better understand how mediation techniques are used by strategy scholars. Finally, implications and limitations of current mediation analysis in strategy research are discussed, and recommendations are provided to strategy scholars examining mediation.
Mediation exists when the relationship between a predictor and an outcome variable occurs through a third variable; this third variable is referred to as a mediating variable. Tests of intricate relationships, including mediating variables, have become more prevalent over the last 10 years in strategic management as researchers have been prompted to explore the processes underlying the relationships they examine (Hitt, Gimeno, & Hoskisson, 1998). While only 20 strategic management articles analyzing mediating effects were published in the *Academy of Management Journal (AMJ)* and the *Strategic Management Journal (SMJ)* between 1986 and 1995, for example, more than twice as many (45 articles) analyzing mediating effects were published in these journals between 1996 and 2005.

Strategy scholars have incorporated mediating effects in a number of diverse research settings. Yli-Renko, Autio, and Sapienza (2001) found that social capital influenced knowledge exploitation through knowledge acquisition. In another example, Baum and Wally (2003) found that organizational structure influenced firm performance through strategic decision speed. Finally, Cho and Pucik (2005) found that innovativeness influenced firm profitability through growth. Taken together, these examples suggest the importance of mediating effects in strategic management research.

Nearly 20 years have passed since the seminal work of Baron and Kenny (1986), henceforth referred to as BK. As such, we attempt to understand how BK has influenced strategic management research through a review of strategy mediation articles. We provide three primary contributions with this article. First, we distinguish between mediating and indirect effect relationships, describe different methods available to test mediating relationships, and discuss the potential implications of this dichotomy with respect to strategy research. Second, we review the strategy research published in the *AMJ* and the *SMJ* between 1986 and 2005 to examine the norms as they pertain to mediation testing in strategy research. Finally, we provide recommendations to strategy researchers for testing mediating effects. Our efforts are not intended to criticize the work of others. Instead, we hope that this review helps scholars to better understand how mediation is used in strategy research.

**MEDIATING VARIABLES**

BK defined mediators as variables that allow an independent variable to influence a dependent variable. Mediation processes occur over time, such...
that the independent variable occurs temporally before the mediating variable and the mediating variable occurs before the dependent variable (Baron & Kenny, 1986; Shrout & Bolger, 2002). The effect of the independent variable on the dependent variable through the mediating variable is referred to as a mediating effect.

Mediation may be either full (also known as “complete”) or partial. In a fully mediated model, the predictor variable, $x$, influences the outcome variable, $y$, only through the mediating variable, $m$. In other words, the entire effect of $x$ on $y$ is transmitted through $m$ (James & Brett, 1984). In a partially mediated model, however, only a portion of the total effect of $x$ on $y$ is due to the mediation by $m$ (Duncan, 1970, 1975; Heise, 1975; Kenny, 1979). In other words, partial mediation suggests that an independent variable influences the dependent variable both directly and indirectly.

We should note that scholars in disciplines such as psychology, sociology, and management have relied on mediation to test hypotheses and often use different terminology. Some scholars, for example, examine intervening variables, which are defined as processes that intervene between a predictor variable and an outcome variable (MacCorquodale & Meehl, 1948; Tolman, 1938; Woodworth, 1928). The effect of the predictor variable on the outcome variable through the intervening variable is referred to as an indirect effect. These terms differ from the concept of mediation, because there is no requirement that the predictor variable have a direct effect on the outcome variable (for a detailed discussion of these terms, see Mathieu & Taylor, 2006).

While it is important to note the differences between these terms, many scholars have used these terms interchangeably (e.g., MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). Although we rely on the concept of mediation in the following sections, we examine mediation as well as other types of intervening effects in our literature review. In other words, we are primarily interested in indirect effects, and requiring a relationship between the independent and dependent variables is not our central concern.

**STATISTICAL TESTS FOR MEDIATING VARIABLES**

MacKinnon et al. (2002) uncovered 14 different analytical approaches that have been used to test for mediating or intervening variables. They reduced these 14 methods into three broad approaches, which we review in the following sections. The first is the causal steps approach, advanced by Judd and Kenny (1981) and BK. The other two approaches, difference in
coefficients and product of coefficients, are tests for intervening variables. We refer readers to MacKinnon et al. (2002) for a complete list and a more comprehensive review of the 14 specific tests.

Causal Steps

The causal steps approach, which was developed by Judd and Kenny (1981) and BK, is a commonly used approach. This approach includes a series of tests, illustrated in Fig. 1. Path $a$ represents the impact of the predictor variable on the mediating variable. Path $b$ represents the impact of the mediating variable on the outcome variable. (The product of $a \times b$ in Fig. 1 constitutes the indirect effect.) Path $c$ represents the impact of the predictor variable on the outcome variable in the unmediated model (the total effect in Fig. 1). Path $c'$ represents the impact of the predictor variable on the outcome variable when the mediating variable is added to the model (the direct effect in Fig. 1) (Baron & Kenny, 1986).

According to BK, testing for mediation consists of four critical steps. First, the predictor variable must influence the outcome variable (path $c$ in Fig. 1). Second, the predictor variable must influence the presumed mediator (Path $a$ in Fig. 1). Third, the mediator must influence the outcome variable while controlling for the predictor variable (Path $b$ in Fig. 1). Finally,
a previously significant relationship between the predictor and outcome variables must be reduced in the presence of the mediator (path $c'$ in Fig. 1).

**Difference in Coefficients**

The second approach, difference in coefficients, is based on comparing the relationship between the predictor and outcome variables before and after adjusting for the mediating variable. Different pairs of coefficients can be compared, including regression coefficients and correlation coefficients (MacKinnon et al., 2002). When using regression coefficients, the difference in coefficients ($c - c'$) from Fig. 1 is used. The difference in coefficients using correlation coefficients is $r_{xy} - r_{xy.m}$, where $r_{xy}$ represents the correlation between the predictor variable and the outcome variable and $r_{xy.m}$ represents the partial correlation between the predictor variable and the outcome variable partialled for the mediating variable, $m$ (MacKinnon et al., 2002). For more details on this approach, see Freedman and Schatzkin (1992), McGuigan and Langholtz (1988), and Olkin and Finn (1995).

**Product of Coefficients**

The third approach, product of coefficients, is based on multiplying the coefficients of the paths in a path model (Alwin & Hauser, 1975; Bollen, 1987; Fox, 1980; Sobel, 1982). The coefficients that are multiplied in this approach are represented by $a$ and $b$ in Fig. 1. The indirect effect of $x$ on $y$ through the mediating variable, $m$, is measured as the product of the $a$ and $b$ paths depicted in Fig. 1 (Preacher & Hayes, 2004). MacKinnon, Warsi, and Dwyer (1995) demonstrated that this measure of the indirect effect is algebraically equal to the difference in coefficients ($c - c'$) for ordinary least-squares regression. This approach tests the significance of the mediating effect by dividing the estimate of the mediating effect, $a \times b$, by its standard error and comparing this value to a standard normal distribution (MacKinnon et al., 2002). The most commonly used product of coefficients formula is that of Sobel (1982), but there are other similar tests (MacKinnon et al., 2002).

To summarize, we have distinguished between the terms “mediate” and “indirect effect” and have discussed various statistical tests to examine potential mediating variables. In the following sections, we assess how
strategy researchers have examined potential mediating relationships. After describing our methodology, we discuss the findings of our literature review and assess those findings in light of more recent methodological advancements published since BK (1986).

**METHODOLOGY**

Our survey of the tests for mediation in strategy research included all strategy articles published in *SMJ* and *AMJ* between 1986 and 2005. Articles from *SMJ* and *AMJ* were examined because they are generally representative of the high-quality strategy research that the field endeavors to conduct. The beginning of the sample window coincides with the publication of BK’s seminal work on mediation. In order to select the sample of articles, we first searched for the terms “mediation,” “mediating,” “mediate,” “intervening,” “indirect effect,” and “intervene” in the text of all articles published in *SMJ* and *AMJ* using Proquest’s ABI/Inform Global Business database. In addition, we searched for articles published in these two journals using “structural equation(s) modeling (SEM)” or “path analysis” during the years of 2003–2005, to include the time period after the Shook, Ketchen Jr., Hult, and Kacmar (2004) study. All articles in Shook et al. (2004) review on SEM were examined individually to determine whether mediation was analyzed. Only articles presenting and discussing some analysis of indirect or mediating effects were included; those making no mention of mediating effects in their analysis or theory were excluded. The resulting set of articles was then further screened to ensure that all *SMJ* articles in our sample were empirical and that all *AMJ* articles fell under the umbrella of strategy.

To distinguish between strategy and non-strategy articles in *AMJ*, we relied upon multiple criteria. First, we turned to the criteria provided by Rumelt, Schendel, and Teece (1994) who suggested that the domain of strategy research is largely concerned with answering four key questions: (1) How do firms behave? (2) Why are firms different? (3) What is the function of, or value added by, the headquarters unit in a multibusiness firm? and (4) What determines the success or failure of the firm in international competition? Articles that addressed any of these questions were included in the sample. Articles examining issues related to corporate governance and strategic leadership were also included (Finkelstein & Hambrick, 1996). Second, articles were evaluated under the Summer et al. (1990) criteria, whereby research investing strategy, environment,
leadership/organization, and performance are included in the strategy domain. To increase the objectivity of the analysis, all of the articles were independently coded by two of the authors. Interrater reliabilities, Intraclass correlation coefficient (ICC) (Shrout & Fleiss, 1979) ranged from .70 to 1, with an average of .86. Disagreements were resolved through discussion between the two coders. This process resulted in 64 articles assessing mediation in strategy research.

**FINDINGS**

From 1986 to 2005, 64 strategy articles published in *AMJ* or *SMJ* hypothesized and tested for mediation. Table 1 summarizes the methods that were used to test mediation. The majority of the strategy papers that we reviewed, 55%, employed structural equations modeling to test for potential mediating effects. In recent years, however, researchers have increasingly relied on SEM to test for mediation. Our content analysis revealed that 26 of the 64 articles in our sample used regression approaches. There were also

<table>
<thead>
<tr>
<th>Number of Studies (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of studies</td>
</tr>
<tr>
<td>Data type</td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Secondary</td>
</tr>
<tr>
<td>Both</td>
</tr>
<tr>
<td>Primary methodological approaches</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>SEM/path analysis</td>
</tr>
<tr>
<td>ANOVA, MANOVA</td>
</tr>
<tr>
<td>Partial correlations</td>
</tr>
<tr>
<td>Mediating effect found</td>
</tr>
<tr>
<td>Supplementary tests of indirect effects</td>
</tr>
<tr>
<td>Sobel</td>
</tr>
<tr>
<td>Goodman</td>
</tr>
<tr>
<td>Bootstrapping</td>
</tr>
</tbody>
</table>

*One study used both Sobel and Goodman. Therefore, the total number of studies testing the significance of the indirect effect is 6.
two articles that used analysis of covariance (ANCOVA) to test for mediation in the articles analyzed. Because ANCOVA was used by only two studies, we will not discuss this further. For a critique of ANCOVA for testing mediators, see Fiske, Kenny, and Taylor (1982). In the following sections, we highlight several observations uncovered through the review.

Dominant Use of BK’s (1986) Four-Step Method

The review of strategy articles revealed that nearly all of the studies in our sample incorporated various aspects of BK’s logic in their methodology whether using regression or SEM in their analysis. BK’s process for testing for mediators includes four steps, as mentioned earlier. In our sample, authors inconsistently completed all four steps. In a small number of cases, authors omitted the first step, which requires a relationship between the predictor variable and the outcome variable.

Although it was rare for authors to exclude the first of BK’s four steps, there are two important reasons why strategy researchers might want to exclude the first step. First, the relationships analyzed in strategy research are frequently distal as opposed to proximal, which makes it more difficult to identify a significant direct effect of the predictor variable on the outcome variable (Shrout & Bolger, 2002). Shrout and Bolger (2002, p. 429) specifically suggested that omitting the first step is logical when the mediator is distal, as opposed to proximal because the distal effect is more apt to be “(a) transmitted through additional links in a causal chain, (b) affected by competing causes, and (c) affected by random factors.”

This distinction between proximal and distal mediators is particularly important in strategy research where distal relationships, longitudinal data, and repeated-measures designs are frequent. Strategy research is concerned with understanding the effects of strategic actions on firm performance where actions and performance outcomes often take place months or even years apart. Relationships between strategic actions and firm performance are complex and distant in time. For example, Krishnan, Miller, and Judge (1997) suggested that top management team (TMT) turnover mediates the relationship between complementarity of the functional backgrounds of the acquired and acquiring firm’s TMTs and postacquisition performance. They measured the dependent variable three years after acquisition to allow time for the complementarity of the executive teams to influence turnover and subsequently firm performance. While distal effects are common in strategy literature, they carry implications for mediation analyses. When the
predictor and outcome variables are temporally separated and the effect is fairly small, it becomes more likely that the effect of \( x \) on \( y \) is transmitted through other intervening variables. This makes it unlikely that a significant direct relationship between \( x \) and \( y \) would exist, as is required in the first step of BK. Forbes and Milliken (1999), for example, acknowledge the difficulty in finding a relationship between board characteristics and firm performance:

> The influence of board demography on firm performance may not be simple and direct, as past studies presume, but rather, complex and indirect. To account for this possibility, researchers must begin to explore more precise ways of studying board demography that account for the role of intervening processes. (p. 490)

Because it is difficult to find a relationship between the predictor and outcome variables when relationships are distal, some scholars assert that mediation analyses may remain useful even when the relationship between the predictor and the outcome variables is not significant (Collins et al., 1998; MacKinnon, 2000; MacKinnon, Krull, & Lockwood, 2000). For example, James and Brett (1984) did not require a direct effect of the predictor on the outcome variable in order to establish mediation.

Second, researchers may also want to skip BK’s first step when a negative mediator is posited because the mediator may cancel out the previous positive relationship between \( x \) and \( y \) (Collins et al., 1998; Frazier, Tix, & Barron, 2004). The mediator may be the root of the failure to find the direct relationship because it may cancel out the direct effect (Collins et al., 1998). MacKinnon et al. (2001) demonstrated the impact of a negative mediator on a positive relationship, showing that the \( x-y \) relationship may be undetected.

For these reasons, requiring a direct relationship between the predictor and outcome variables may impede research in strategic management. The difficulty in finding distal relationships may discourage researchers from continuing mediation analysis if the relationship between the predictor and the outcome is not found to be statistically significant per BK’s first step. One study in our analysis did not find mediation because they did not find significance in the first step (Green, Welsh, & Dehler, 2003). Green et al. (2003, p. 429) commented that “this finding [that \( x \) was not significantly related to \( y \)] obviously precludes the possibility of advocacy’s mediating the relationships between project characteristics and project terminations.”

Because strategy researchers may have difficulty finding significance in the first step when testing distal relationships, there may be several mediation relationships in strategy that we have no knowledge of due to the file drawer problem (Rosenthal, 1979). Rosenthal (1979) believed that there is a
publication bias that occurs when the probability that a study gets published is based on the statistical significance of the results. One reason why we may find so few published strategy articles with mediation is, perhaps, that many fail to establish the conditions necessary to proceed to BK’s second step.

**Issues Regarding Statistical Power**

Our review of the strategy literature also highlighted the need to incorporate statistical power in tests of mediation. Statistical power is defined as the ability of a technique to detect relationships present in the data (Vogt, 1993). Cohen (1988, 1992) suggested that researchers should aim to achieve a power level of .80. In this section, we discuss several issues that may affect power when testing for mediation, including the statistical test, collinearity between the predictor and mediator variables, reliability of the mediator, and sample size.

Recent research has questioned the statistical power associated with the BK approach, which represents the most dominantly used approach to test for mediation in the sample (MacKinnon et al., 2002). In a simulation study comparing methods for testing intervening effects, MacKinnon et al. (2002) found that the BK and Judd and Kenny (1981) methods had very low power for detecting small and medium effect sizes. MacKinnon et al. (2002, p. 96) suggested that “studies that use the causal steps methods described by Kenny and colleagues are the most likely to miss real effects.” In their simulation, MacKinnon et al. (2002) found that BK only had power ranging from 0.0040 to 0.1060 for small effect sizes ranging in sample size from 50 to 1,000 respectively. Other methods, such as Freedman and Schatzkin’s (1992) difference in coefficient method, have power as high as .99 for small effect sizes with a sample size of 1,000 (MacKinnon et al., 2002). Therefore, reliance on the BK method is concerning because strategy research tends to have small effect sizes (Hitt, Boyd, & Li, 2004), which puts such studies at risk of missing effects.

Aside from the fact that the BK approach has low power to detect small effect sizes like those common in strategic management research, there are additional factors that may further reduce power. For instance, the relationship between the mediator, predictor, and outcome variables can influence the power of the test of mediation (Frazier et al., 2004). The power of the test of the mediator–outcome relationship (path b in Fig. 1) and the predictor–outcome relationship when controlling for the mediator variable (path c’ in Fig. 1) decreases as the relationship between the predictor
variable and the mediator variable (path $a$ in Fig. 1) increases (Frazier et al., 2004; Kenny, Kashy, & Bolger, 1998). In other words, as more variance in the mediator is explained by the predictor variable, there is less variance in the mediator to predict the outcome variable. Therefore, as the relationship between the predictor variable and the mediator variable increases (path $a$ in Fig. 1 gets larger), researchers must rely on larger samples in order to have enough power to test the effects of the other two paths in the model (Frazier et al., 2004; Kenny et al., 1998).

Kenny et al. (1998) provide a formula to demonstrate how power diminishes as the correlation between the predictor variable and mediator grows. They found that power is reduced in mediation analyses, making the effective sample size equal to $N(1-r_{xi}^2)$, where $N$ is the sample size and $r_{xi}$ is the predictor ($x$) to mediator ($m$) correlation. According to their framework, a sample size of 100 and a correlation of .80 between the predictor and the mediator ($r_{xi} = .80$) implies an effective sample size of 36 (e.g., $N(1 - r_{xi}^2) = 100(.36) = 36$). Because power is often an issue in the BK causal steps regression approach, determining the necessary sample size represents a step toward addressing the limitations of that approach.

The reliability of the mediator may also influence the power of mediation tests (Frazier et al., 2004). Reliability is important in strategy research, as demonstrated by the fact that 49 of the 64 articles reviewed used some method of primary data that often used surveys where reliability was computed and reported. Reliability is frequently measured with Cronbach’s alpha (Cronbach, 1951), using the rule of .70 or greater reliability as acceptable (Nunnally, 1978).

When reliability is low, the effect of the mediator variable on the outcome variable (path $b$ in Fig. 1) is underestimated and the effect of the predictor variable on the outcome variable (path $c'$ in Fig. 1) is overestimated (Baron & Kenny, 1986; Judd & Kenny, 1981; Kenny et al., 1998). Therefore, statistical analyses such as multiple regression, which ignore measurement error, may underestimate the mediation effects (Frazier et al., 2004). Hoyle and Kenny (1999) and Hoyle and Robinson (2003) provided a formula for estimating the effect of low reliability on tests of mediation.

Because of the power problems that can arise due to low reliability and collinearity, several researchers have offered heuristics to help ensure tests of mediation have adequate power. Hoyle and Kenny (1999) suggested that when the mediator is highly reliable ($\alpha = .90$), a sample size of 100 will yield the necessary power. However, they suggested a sample size of at least 200 when mediators have moderate reliability ($\alpha = .70$). Additionally, Kenny et al. (1998) stated that researchers testing for mediation often need a sample
size greater than the average study in their field because of low reliability and high collinearity between the predictor and mediator variables, which lowers power.

Although Hoyle and Kenny (1999) and Kenny et al. (1998) offered these heuristics, none of the studies that we reviewed mentioned whether they considered these or other guidelines pertaining to statistical power. Reliabilities of the measures used were most often between .70 and .80. While the average sample size did not appear to be a problem in the majority of studies, effect sizes in strategy research appear to be fairly small, warranting other methods than BK to be used.

**Testing for the Significance of the Indirect Effect**

Traditionally, scholars have held that full mediation is established and significant only when the predictor–outcome effect goes from “significant” to “not significant” once the mediator is added to the model (Holmbeck, 2002). However, Holmbeck (2002, p. 88) pointed out that this measurement might not be precise enough because “a drop in significance to non-significance may occur … when a regression coefficient drops from .28 to .27 but not when it drops from .75 to .35.” In other words, it is possible to have the predictor–outcome relationship drop from significant to not significant when accounting for the mediator even though there is no significant mediation, or for a mediating effect to be present when the predictor–outcome relationship continues to be statistically significant even after adding the mediator into the model. Therefore, Holmbeck (2002) concluded that researchers must test for the significance of the mediating effect because the results of studies that do not test the significance of the mediating effect may be spurious.

Frazier et al. (2004, p. 128) corroborated Holmbeck’s argument and stated that “it is not enough to show that the relation between the predictor and outcome is smaller or no longer significant when the mediator is added to the model.” Instead, a method for testing the significance of the mediating effect should be used. Preacher and Hayes (2004) also argued for formally testing the indirect effect. They agreed with Holmbeck (2002) that without testing the indirect effect, researchers are more likely to make Type I errors if the addition of the mediator to the model causes a very small change such that a statistically significant relationship between $x$ and $y$ becomes non-significant. This is true especially when the sample size is large and even small regression weights
are statistically significant (Preacher & Hayes, 2004). Researchers are also more likely to make Type II errors when there is a large change in the relationship between \( x \) and \( y \), but there is no observed drop in significance. Therefore, failing to test for the significance of the indirect effect may result in spurious findings.

There are many approaches for testing the significance of mediation effects, such as the Sobel’s first-order solution, the Goodman unbiased solution, and the Freedman and Schatzkin method (MacKinnon et al., 2002). In order to calculate the indirect effect, the weights for paths \( a \) and \( b \) as well as their respective standard errors are required. Sobel (1982) is a commonly used method to test for the indirect effect. In order to perform this test, the indirect effect, \( ab \), is divided by the standard error of \( ab \), \( s_{ab} \) which is defined as: 

\[
s_{ab} = \sqrt{b^2s_a^2 + a^2s_b^2 + s_a^2s_b^2}.
\]

The formula yields a ratio that is compared with the critical value from the standard normal distribution to determine whether the indirect effect is statistically significant (Preacher & Hayes, 2004). Preacher and Leonardelli (2003) have even posted a web page with a Sobel calculation tool for mediation tests, which can be found at: http://www.unc.edu/~preacher/sobel/sobel.htm

Because of the simplicity of testing the indirect effect, the benefits of testing for the indirect effect far outweigh the costs. However, only 10% of strategy studies in our review used any of these tests. Testing for statistically significant indirect effects in the articles analyzed varied based on whether the authors used regression or SEM for mediation analysis. Only five articles that we analyzed tested the significance of the indirect effect. Of those five articles, we identified four that used Sobel’s (1982) test in addition to the BK approach. For example, after analyzing MANCOVA results, Sapienza and Korsgaard (1996) evaluated whether perceptions of procedural justice mediate the effect of timely feedback on entrepreneur–investor relations and then estimated the significance of the indirect effect using Sobel. Although the use of the Sobel test was discussed in BK’s work, 90% of the studies in our sample did not mention the significance of the indirect effect.

To summarize, our literature review of strategy articles revealed that the BK approach is overwhelmingly the most common method used to test for mediation. While strategy scholars have displayed a mastery of the BK method, several issues should be considered, including whether BK’s first step should always be followed, issues regarding statistical power, and the importance of testing for the statistical significance of the indirect effect. Therefore, in the next section, recommendations are provided for strategy scholars examining mediation.
RECOMMENDATIONS

Mediation analysis is critical to strategy research because it detects the mechanism by which a predictor variable influences an outcome variable. As such, the use of mediation has the potential to explain complex relationships in the strategic management literature. Therefore, we provide a number of recommendations for mediation analysis.

Tips from the Trenches for Beginners Using BK

First, we offer some fundamental tips for beginners learning how to run mediated regression using BK’s four-step method. When beginning mediation analysis, it is a good idea to look at the correlation matrix in order to understand the magnitude of the correlations between the variables in question. This will help you understand the nature of the bivariate relationship between the $x$, $m$, and $y$ variables. For example, if the $x \rightarrow y$ relationship is weaker than the $m \rightarrow y$ relationship, this provides initial support for mediation. This is because the mediator is more proximal to the dependent variable and thus should have a stronger relationship with it.

It is also advisable to look at the signs of the correlations between the $x$, $m$, and $y$ variables. If the $x \rightarrow m$ and the $m \rightarrow y$ relationships have opposite signs (i.e., one is negative and one is positive), this may explain a situation where there is no significant relationship between the $x$ and $y$ variables. Because the two relationships are in opposite directions, they cancel out and produce no obvious relationship between $x$ and $y$. This could then be a good reason to skip Step 1 of BK since a non-significant relationship between $x$ and $y$ should not preclude you from completing the other BK steps (Shrout & Bolger, 2002).

When you are ready to run mediated regression, be sure to input the variables into three regressions (assuming you will complete all four steps of BK). To illustrate, below are the steps of how variables are entered into each equation:

Step 1: In the first regression, regress the outcome variable ($y$) on the predictor variable ($x$) to establish that there is a relationship between $x$ and $y$.

Step 2: In the second regression, regress the mediator ($m$) on the predictor variable ($x$) to estimate the relationship between $x$ and $m$. 
Step 3: In the third regression, regress the outcome variable \( (y) \) on both the predictor \( (x) \) and the mediator \( (m) \) simultaneously to assess the relationship between \( x \) and \( y \) while controlling for \( m \).

Step 4 of the BK method is where you interpret whether the results show full or partial mediation, or any mediation for that matter. This step requires no additional computation; it relies on interpretation of the first three steps. Now, we will interpret a simple example of mediated regression.

Let us assume that we are testing the hypothesis that the relationship between TMT heterogeneity and firm performance is mediated by cognitive conflict. (Please see Table 2 for the sample data.) Here, we can see in Step 1 that TMT heterogeneity significantly predicts firm performance \( (p<.01) \). In Step 2, TMT heterogeneity is a significant predictor of cognitive conflict \( (p<.01) \). However, if TMT heterogeneity was not significantly related to cognitive conflict, there would be no support for mediation. In Step 3, TMT heterogeneity is no longer a significant predictor of firm performance when cognitive conflict is also entered, and cognitive conflict is a significant predictor of firm performance \( (p<.01) \). Therefore, this is an example of full mediation, because the significant effect of TMT heterogeneity as a predictor of firm performance is no longer significant when cognitive conflict is in the model. The effect of TMT heterogeneity on firm performance is transmitted through the mediator, cognitive conflict.

\[
\begin{array}{ccccccc}
\text{Table 2.} & \text{Interpretation of Sample-Mediated Regression Results.} \\
\hline
\text{Variable} & b & \beta & \text{SE} & t & R^2 & F \\
\hline
\text{Step 1} & & & & & & \\
\text{DV = Firm performance} & \text{1. TMT heterogeneity} & .44^{**} & .22^{**} & .15 & 2.97 & .05 & 8.81^{**} \\
\text{Step 2} & & & & & & \\
\text{DV = Cognitive conflict} & \text{1. TMT heterogeneity} & .37^{**} & .32^{**} & .08 & 4.40 & .10 & 19.36^{**} \\
\text{Step 3} & & & & & & \\
\text{DV = Performance} & \text{1. TMT heterogeneity} & .20 & .10 & .15 & 1.39 & & \\
\text{} & \text{2. Cognitive conflict} & .65^{**} & .37^{**} & .13 & 5.00 & .17 & 17.54^{*} \\
\hline
\end{array}
\]

* \( p < .05 \).
** \( p < .01 \).
If the TMT heterogeneity variable had been reduced in significance (down to the .05 level, for example) in Step 3, there would be evidence of partial mediation. Here, we would say this is a partially mediated relationship, because the significance of TMT heterogeneity as a predictor of firm performance is reduced when cognitive conflict is in the model. This means that some of the effect of TMT heterogeneity on firm performance is transmitted through cognitive conflict, but not all. Finally, if the statistical significance of TMT heterogeneity had remained unchanged (at the $p < .01$ level) in Step 3 or cognitive conflict was not significantly related to firm performance controlling for TMT heterogeneity in Step 3, then there would be no evidence of mediation at all.

If a mediating effect is identified as a result of BK Step 4, the final step in the test of mediation should be testing the statistical significance of the indirect effect. Please see Table 3 for an example showing a manual calculation of the Sobel formula to test the significance of the indirect effect using the data from the paths $a$ and $b$ standard errors and coefficients in the TMT heterogeneity mediation example from Table 2. The result of the Sobel formula is also interpreted in Table 3, showing that the indirect effect is statistically significant.

**Table 3.** Example Test of the Sobel Equation to Test the Statistical Significance of the Indirect Effect.

<table>
<thead>
<tr>
<th>$Z = ab / \sqrt{S_a^2 S_b^2 + b^2 S_a^2 + a^2 S_b^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a = $ Path $a$ from Fig. 1</td>
</tr>
<tr>
<td>$b = $ Path $b$ from Fig. 1</td>
</tr>
<tr>
<td>$S_a = $ standard error of path $a$</td>
</tr>
<tr>
<td>$S_b = $ standard error of path $b$</td>
</tr>
</tbody>
</table>

Sobel (1980) test for indirect effect

| $a$   | 0.37 |
| $b$   | 0.63 |
| $S_a$ | 0.08 |
| $S_b$ | 0.13 |

Numerator $a \times b$ (indirect effect)

| 0.23 |

Denominator

| 0.07 |

$Z$-score for indirect effect

| 3.28 |

The $Z$-score of 3.28 is greater than the $Z$ critical value of 1.96, which means the $Z$ score for the indirect effect is statistically significant ($p < .01$).
Tips for Those Who are More Familiar with BK

While the BK approach for detecting and testing mediation remains one of the most influential approaches, strategic management researchers should consider more recent advancements in mediation testing that have been published since BK’s seminal work. Therefore, we offer some more advanced recommendations, highlighted in Table 4, for researchers who already understand the mechanics of the BK method. First, when the predictor–outcome relationship is distal, we recommend considering whether BK’s first step is relevant (Collins et al., 1998; Shrout & Bolger, 2002). Our recommendations follow the advice of Shrout and Bolger (2002, p. 430) who write, “Relaxing the requirement that $x \rightarrow y$ be statistically significant before going on to study mediation is likely to be especially important for … researchers who track long-term processes.” With the prevalence of distal relationships in strategy research, scholars should think about the conceptual importance of the direct effect. If the direct effect is not conceptually important, the indirect effect through the intervening variable is just as theoretically valuable, and the researcher may omit the first step of BK. Relaxing this assumption may reduce rejection of studies that may have indirect effects, as well as the file drawer problem whereby researchers stop the research when they do not find a significant relationship in BK’s first step.

Second, researchers should consider those elements that may reduce power to detect a mediating effect. The relationship between the mediator variable and the predictor variable can affect the power of the test of

Table 4. Recommendations for Strategy Scholars.

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Suggested Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider the need to skip Baron and</td>
<td>Shrout and Bolger (2002), Collins, Graham, and Flaherty (1998), Frazier et al.</td>
</tr>
<tr>
<td>Kenny’s Step 1</td>
<td>(2004)</td>
</tr>
<tr>
<td>Evaluate the reliabilities of the</td>
<td>Hoyle and Kenny (1999), Frazier et al. (2004), Kenny et al. (1998), MacKinnon</td>
</tr>
<tr>
<td>variables and the implications on</td>
<td>et al. (2002)</td>
</tr>
<tr>
<td>power</td>
<td>Shaver (2005)</td>
</tr>
<tr>
<td>Evaluate the possibility of error term</td>
<td>Holmbeck (2002), MacKinnon et al. (2002), Preacher and Hayes (2004), Holmbeck</td>
</tr>
<tr>
<td>correlation</td>
<td>(2002)</td>
</tr>
<tr>
<td>Test the significance of the indirect</td>
<td>James et al. (2006), Hoyle and Smith (2004), Bollen (1987), Brown (1997)</td>
</tr>
<tr>
<td>effect</td>
<td></td>
</tr>
</tbody>
</table>
mediation (Frazier et al., 2004). Therefore, scholars need to consider the collinearity between these two variables and its effect on power. We recommend Kenny et al.’s (1998) formula as way to evaluate this effect.

In addition, an unreliable mediator underestimates the mediator’s effect on the outcome variable (Baron & Kenny, 1986; Judd & Kenny, 1981). Due to the problems that can arise with low reliability, Hoyle and Robinson (2003) suggested using a measurement instrument with a reliability of .90 or higher, and Hoyle and Kenny (1999) recommended sample sizes of 100 for highly reliable mediators or 200 for moderately reliable mediators. Because none of the articles reviewed discussed power, this could possibly indicate that strategy researchers, on the whole, pay little attention to power with regard to mediation. Thus, we urge scholars to consider power, and we suggest Cohen’s (1992) work which provides a concise benchmark for assessing sample size given desired power, estimated effect size, and alpha as a starting point.

Third, we urge researchers to test the statistical significance of the indirect effect. Similar to MacKinnon et al. (2002), we found that most scholars did not test the statistical significance of the mediator. Given the power issues associated with both mediation and strategic management research in general, researchers should test the significance of the indirect effect (Hitt et al., 2004; MacKinnon et al., 2002). In extant research, Sobel’s (1982) test has been the most popular approach. (For a description of how SPSS and SAS calculate the Sobel test, see Preacher & Hayes, 2004.) However, there are also other formulas to test for the statistical significance of the indirect effect. In MacKinnon’s review, the difference in coefficients method by Freedman and Schatzkin (1992) had the highest power of all 14 methods tested.

Finally, we direct those testing mediation using multiple regression to some helpful tools developed recently by mediation experts. Frazier et al. (2004) includes an appendix with a helpful checklist for mediation analysis including questions such as “Was the predictor significantly related to the outcome? If not, was there a convincing rationale for examining mediation?” “What is the ‘effective sample size’ given the correlation between the predictor and the mediator?” “Was power mentioned either as an a priori consideration or as a limitation?” “Was unreliability in the mediators (e.g., $\alpha < .70$) addressed through tests that estimate the effects of unreliability...?” , and “Was the significance of the mediation effect formally tested?” In addition to this checklist, both Kenny (http://davidakenny.net/cm/mediate.htm) and MacKinnon (http://www.public.asu.edu/~davidpm/ripl/mediate.htm) have web sites on mediation.
We also suggest SEM as an alternative to the BK causal steps approach. We observed that over half of the studies testing mediation use SEM (e.g., Robins, Tallman, & Fladmoe-Lindquist, 2002; Simonin, 1999; Tippins & Sohi, 2003). We agree that this approach may be helpful because SEM allows the researcher to look at all the data simultaneously without necessarily making causal inferences. SEM also has the capacity to compare and contrast alternative models to identify the most likely causal direction. This methodology, combined with good theoretical rationale, can provide authors with some insight into the nature of the relationships present in the data even in non-lagged data.

James, Mulaik, and Brett (2006, p. 243) strongly urged that researchers “add tests of alternative causal models to basic mediation analysis” and suggest this as one of the key advantages of SEM. Because of the difficulty in establishing temporal priority in non-experimental research, studies that have improperly specified models are at risk of testing incorrect models (Stone-Romero & Rosopa, 2004). In addition, competing models may often fit the data similarly (Stelzl, 1986). Alternative models may be tested to lessen this risk. However, our review indicated only limited use of alternative models. Clearly, SEM can never overcome a flawed research design or substitute for having the appropriate lags between variables. Nonetheless, this technique can potentially help researchers better understand the nature of the relationships between variables and inform researchers of the viability of each model (Jermier & Schriesheim, 1978). Because of this, we recommend that researchers consider using SEM to test for mediating effects.

Tests of the significance of the indirect effect are also available in most SEM programs (Frazier et al., 2004). (See Brown (1997) for a description of testing mediation models in LISREL.) Because of SEM’s sophistication and flexibility, many researchers, including BK, have described SEM as the “most efficient and least problematic means of testing mediation” (Hoyle & Smith, 1994, p. 438).

Shaver (2005) suggested that an additional advantage to using SEM is that it can help address problems resulting from correlated error terms. Measurement error may have a detrimental effect on the interpretation of mediation analysis because it may induce the error terms to correlate (Shaver, 2005). When a variable affects both \( m \) and \( y \) in the same way, it causes correlation of the error terms in the second and third steps of BK. In a review of management articles, Shaver (2005) found that the error terms could correlate due to missing variables, measurement error, and truly random effects. Correlation becomes a problem because the coefficient
estimates are inconsistent to the extent that they are correlated. Hence, Shaver (2005) recommended simultaneous equation problem techniques such as SEM to deal with these issues.

Like other methods, SEM is not perfect. SEM suffers from the problem of omitted variables that can affect the models being tested (Cliff, 1983; Freedman, 1987; Tomarken & Waller, 2005). SEM has also been criticized for encouraging researchers to focus on global fit of the model at the expense of lower-order model components that may affect the model (Tomarken & Waller, 2003). Finally, researchers have also noted that using SEM to conduct multilevel analysis can become unwieldy as the model becomes more complex (Chen, Bliese, & Mathieu, 2005). Despite these limitations, SEM provides a number of unique and important benefits which researchers should consider (Baron & Kenny, 1986; Hoyle & Smith, 1994; Judd & Kenny, 1981). (For a review and suggestions on how to best use SEM in strategic management research, see Shook et al., 2004).

CONCLUSION

Strategic management is a relatively young yet maturing academic discipline. As such, many propose that it is necessary to examine the state of the field’s methodological sophistication and maturity. The purpose of this study has been to assess how mediation analyses are used in strategy research and to assess this methodology in light of methodological advancements published in the last 20 years since BK’s seminal work. In order to achieve this end, we examine articles published in AMJ and SMJ from 1986 to 2005. We find that tests of mediation in published strategy research are relatively infrequent. The results of our review reveal that while BK remains the approach most commonly cited by strategy researchers, there is variance in how BK’s approach is implemented. Specifically, we find that researchers tend to rely heavily on BK’s step. In addition, they infrequently omit the first step, discuss the implications of power, or test the significance of the indirect effect.

Drawing on extant literature on mediation, we suggest that the BK approach should be used with caution due to its low power (MacKinnon et al., 2002) and the distal relationships associated with strategy research (Shrout & Bolger, 2002). As a result, we recommend that strategy researchers open a dialogue about how best to test for mediating relationships in strategy research. It is our hope that this study will spark
discussion concerning how best to test for mediating effects in strategy research.

REFERENCES


Mediation in Strategic Management Research


USING POLICY CAPTURING TO UNDERSTAND STRATEGIC DECISIONS – CONCEPTS AND A MERGERS AND ACQUISITIONS APPLICATION

Amy L. Pablo

ABSTRACT

This chapter outlines purpose, procedure, benefits and limitations of the policy capturing methodology. It further presents an example of use of the methodology.

INTRODUCTION

A common focus in strategic management research is on attempting to explain why certain strategic actions were taken. Indeed, in the last two decades, organizational and strategic management scholars have begun to conceptualize strategy in terms of decision-making processes. Strategy process research is primarily focused on the actions that lead to and support strategy (Huff & Reger, 1987). It is through this cognitive perspective that we have come to understand how strategic decision occasions are recognized, diagnosed, and acted upon. Strategy and other organizational researchers want to be able to develop and test theory about what elements
of information are influential in decisions that have a central impact on organizational outcomes.

Of course, we could just ask an organization’s management and other members, what happened and why, but can it really be that easy? Many researchers have used retrospective reports to carry out their work because they are theoretically an easy source of data. But even if they are available, is this the best source? This type of data is a popular tool for learning about the past (Miller, Cardinal, & Glick, 1997), but it like other survey methods has been shown to be problematic. As demonstrated by Golden’s (1992) work, inaccuracies due to faulty memory, cognitive biases, retrospective justification, time lapse, and event infrequency calls into question the merit of using retrospective reports for understanding strategic decisions.

Another method frequently used to understand human judgment is that of policy capturing. This is a regression-based technique founded on social judgment theory from social psychology and is especially valuable for understanding decisions when multiple informational cues are available to decision makers (Stumpf & London, 1981), and is invaluable to researchers whose goal is learning which pieces of information are most influential in determining decisions. An excellent primer on this research method was published by Aiman-Smith, Scullen, and Barr (2002), and should be referenced for essential information on study design, study execution, analysis, interpretation, and reporting of results.

The purpose of this chapter is to enlighten researchers interested in or needing to understand factors affecting decisions about strategic initiatives for their firms. The methodology is valuable in testing extant theory, comparing different theoretical perspectives, assessing usefulness of new theories in explaining decision processes, and more. This method enables researchers to capture managers’ actual “theories-in-use” which are revealed in the decisions they make when presented with certain informational cues (Hitt & Tyler, 1991; Pablo, 1995). In this chapter, we will discuss some basics of policy capturing, and as an illustration, present some prior work from the mergers and acquisitions (M&A) context and how decisions are made about the integration question so important to outcomes from this organizational strategy.

POLICY CAPTURING OVERVIEW

A “policy” is a plan designed to influence and determine decisions, actions, and other matters; it is a course of action, guiding principle, or procedure
considered to be expedient, prudent, or advantageous (American Heritage Dictionary, 2000, p. 1014). One of the oldest and most widely used techniques for gaining an understanding of individuals' decision processes is policy capturing (e.g., Christal, 1968; Hoffman, 1960; Slovic & Lichtenstein, 1971). The purpose of this technique is to infer the implicit "policy", or model guiding an individual’s decision process by observing the relationships between decision criteria and the decisions of the individual (Hobson & Gibson, 1981; Hoffman, 1960; Stumpf & London, 1981). Research using this technique centers around one broad question: “What is the decision maker doing with the information available to him?” (Slovic & Lichtenstein, 1971, p. 17).

Policy capturing has been used to investigate a range of managerial decisions including promotion decisions (Stumpf & London, 1981), performance appraisal (Hobson, Mendel, & Gibson, 1981; Taylor & Wilsted, 1974; Zedeck & Kafry, 1977), and the selection of acquisition candidates (Hitt & Tyler, 1991; Stahl & Zimmerer, 1984) among others. It has also been used to explore what lies behind decisions at other organizational levels including employees' decisions to trust after an acquisition (Stahl, Chua, & Pablo, 2007). Thus, one of the benefits of policy capturing is that it allows researchers to work backward from organization members' decisions to either evaluate existing theory or to develop new theory in a variety of decision-making domains.

“Values are the instruments through which we select from more information than we can handle ... the simplified constructed situations in which we can act” (Argyris & Schön, 1974, p. 162). As such, values drive the processes of information gathering and interpretation that form the cognitive framework within which managers make strategic decisions impacting on organizations. Increasingly, management scholars are focusing on understanding the links among managers’ cognitions, actions, and organizational performance (e.g., Thomas, Clark, & Gioia, 1993). Yet managers’ understanding of those relationships may be limited, resulting in suboptimal conditions for learning and effectiveness in strategic decision situations (Pablo, 1995).

Procedure

Policy capturing uses regression analysis, or a correlational paradigm (see Slovic & Lichtenstein, 1971, for a review of techniques for studying information processing in judgment) to simulate individuals’ processing of
information in making judgments (Billings & Marcus, 1983; Slovic & Lichtenstein, 1971; Slovic, Fischhoff, & Lichtenstein, 1977). That is, statistical models are developed that describe how to weight and combine the variables of interest in order to accurately reproduce the individual’s decisions.

The methodology is characterized by the presentation to decision makers of a series of decision situations varying in terms of the scores on a number of decision criteria (the independent variables). Decision criteria are determined theoretically and the policy capturing approach tests the manner in which decision makers appear to use the criteria specified by the theories. The factors suggested as determinants (the independent variables) of the decision ultimately made are used as decision criteria around which a set of decision scenarios is constructed to simulate the decision process. Decision makers are instructed to review each decision situation and then assign an overall, global rating (a unidimensional dependent variable) that best summarizes or represents their judgment in the situation given the information available. Using this data on the independent and dependent variables, multiple regression techniques are then used to estimate a statistical equation that represents or “captures” each decision maker’s implicit decision policy, as inferred from the relationship between the predictor and criterion variables.

The equation describes the consistency or reliability with which the decision maker used the decision criteria in making the overall judgments (represented by the squared multiple correlation coefficient ($R^2$), the consistency index for each individual), and the relative importance of each decision criterion in determining the overall judgment (represented by the calculated vector of multiple correlation coefficients or beta weights). That is, $R^2$ represents the extent to which an individual’s decisions may be reproduced by the linear additive combination of the predictor variables. The beta weights represent the relative degrees of importance of the predictor variables reflected in the individual’s decisions.

Once regression weights have been derived, they primarily serve as a vector of data points that describe the subjects in the study. That is, the regression weights can be compared, clustered, averaged, or treated in other ways as new data (Anderson, 1969; Stumpf & London, 1981). The resultant policy equation, therefore, explicitly describes the relationship between the characteristics of the decision situation as represented by the decision criteria, and the individual decision maker’s global assessment of the situation.

A number of researchers (e.g., Hoffman, 1960; Beach, 1990; Christal, 1968; Dawes & Corrigan, 1974; Dawes, 1979) have found that a purely
additive linear model adequately describes the relationship between an individual’s decisions and the criteria used to arrive at those decisions. However, more complex equations can be used to capture curvilinear relationships (e.g., cues are raised to a power in the policy equation) or configural relationships (e.g., cross-product terms, or interactions, are incorporated into the policy equation) (Christal, 1968; Slovic & Lichtenstein, 1971). Finally, structured interviews can be conducted with policy capturing respondents to add richness via a detailed knowledge of context (McGrath, 1982) to interpretation of the quantitative data gathered. The strengths of each method complement the weaknesses of the other, allowing for greater understanding than could be achieved in a monomethod study (Jick, 1979; Martin, 1984; Sieber, 1978).

In the section below, we present an example of a policy capturing study done to learn how key decision makers in acquisitive organizations use variables suggested by extant theories in making decisions about acquisition integration (Pablo, 1995).

Example

Despite the historically poor performance record of this strategic alternative, M&A experts say 2006 is likely to be a record year for M&A activity, swamping the $3.327 trillion in global volume and $1.525 trillion in U.S. volume posted in 2000 (money.cnn.com, 20 December 2005). As such, acquisitions represent a strategic activity having a major impact on the economy and the organizational landscape. Thus, reviewing Pablo’s (1995) policy capturing work on decision maker insight into M&A integration here is both timely and valid.

Integration is central to the process of creating value through acquisitions and the integration design that is chosen is inextricably linked to acquisition performance (Cannella & Hambrick, 1993; Datta & Grant, 1990; Shanley, 1987; Shrivastava, 1986). It is the mechanism through which an acquisition’s synergistic potential is realized by fine-tuning the ways in which two organizations’ operations, systems, and decision frames of reference interact. As such, integration design decisions are those involving choices about the degree of post-acquisition change in an acquired organization’s technical, administrative, and cultural configuration (Pablo, 1994). The central importance of this concept is reflected in the volume of recent attention to it in the literature (e.g., Schweiger, 2002; Fubini, Price, & Zollo, 2006).
As the individuals primarily responsible for exercising governance over newly added business units, the acquirer’s top managers play a critical part in determining the direction of the combined business. That is, much as it is the role of the leader to act as an organization designer (Senge, 1990), it is the role of acquiring management to act as designer of a plan for integrating an acquired business into a parent organization. In this role, they must identify and communicate the core values by which the acquisition process will be implemented and provide the governing ideas that will translate into the policies, structures, and practices by which the newly combined organization will be managed. As Haspeslagh and Jemison (1991) note, such communication is particularly critical during the post-acquisition transition phase, since at this stage operating responsibility for interface management between the two companies is handed off to a new individual or team of managers appointed to execute the vision of the integration designers. Thus, integration designers must be capable of selling the “logic and timing” of their integration plan (Haspeslagh & Jemison, 1991, p. 202) and therefore must be clear on the values they use as guiding principles and the mental models they produce.

Fubini et al. (2006) have reiterated these points in their recent work highlighting the role of top management in thwarting problems that plague many mergers: poorly defined strategic fit, under-resourcing of the integration team, and lack of attention from senior management. This work defines a leadership role for CEOs to help meet aspirations for the new company. These foundations of corporate health include creating the new company at the top, communicating the corporate story, establishing a new performance culture challenge, championing external stakeholders, and fostering momentum and learning. The recommendations laid out here confirm and provide further support for the importance of the leadership role previously identified by Sitkin and Pablo (2004) in their work on leadership in M&As. They specify the importance of six essential dimensions to effective leadership, each of which has a specific effect on followers: personal leadership fosters loyalty, relational leadership engenders a sense of trust and justice, contextual leadership helps to build community, inspirational leadership encourages higher aspirations, supportive leadership forges an internalized sense of self-discipline, and stewardship raises an internalized sense of responsibility.

Mergers and acquisitions are relatively sporadic events for most organizations (Jemison & Sitkin, 1986) and are even more so for the individuals involved in the decision-making processes that surround them (Haspeslagh & Jemison, 1991). Thus, the insights that can be gained from
each decision-making experience are invaluable for the learning process and for application in future acquisitions. Such learning, however, implies that managers have an understanding of their acquisition-related “theories-in-use” (Argyris & Schön, 1974) and how those relate to acquisition outcomes. While in M&A target evaluation, the prescriptions of analytical strategic planning models provide formal decision rules that serve to increase managers’ insight into the nature and consequences of the values guiding their candidate choices (cf., Nisbett & Wilson, 1977), there is much less reason to be so sanguine about the even more critical element of the acquisition process. The same degree of insight is unlikely in acquisition integration decision-making as it is an unstructured, ambiguous task for which little formal training or widely accepted normative model exists. In such situations, tacit subjective ideologies and values used to simplify and structure decision-making will dominate (Beyer, 1981).

Five key features of acquisition situations have been found to enter into acquiring managers’ decision models in developing integration design strategies (Pablo, 1994), and the extent to which each of these is influential reflects the decision-related values the integration designer uses to “frame” the task (Beach, 1990). Managers who value the coordination possible with strong linkages between the firms’ activity chains will emphasize the need to share strategically critical skills and resources (strategic needs). Similarly, integration designers who are primarily concerned with exploiting the benefits to be derived from differentiation and diversity will stress the need to preserve unique organizational capabilities within the acquired firm (organizational needs). For executives who see culture as an important “internal variable” (Smircich, 1983) influencing organizational identity, commitment, and stability, their own organization’s tolerance for differences in values, philosophies, and beliefs (multiculturalism) will be salient in integration planning. Managers who subscribe to the notion of control through organizational dominance will be concerned with the relative sizes of the two organizations (power differential), and for executives to whom having the acquiree’s willing collaboration is pivotal, the extent to which acquirer and target share compatible visions about post-acquisition roles and goals (compatibility of acquisition visions) will be key to developing an integration plan.

This study examines the extent to which 56 top-level managers understand and thus, ultimately can articulate and communicate the tacit values and mental models they use to make integration design decisions. By comparing these behaviorally demonstrated theories-in-use to “espoused theories” that are communicated, expressing preferences for certain types of decision
behavior, we can determine whether discrepancies between the two exist and where the greatest opportunities for organizational learning about acquisitions exist. As Fubini et al. (2006) argue, only when this type of “deep learning” (Senge, 1990) is accomplished organizations will be capable of refining their own ability to use acquisitions as a successful strategy for corporate growth and renewal.

In this research it was hypothesized that managers would have little insight into their integration decision-making models. Based on the characteristics of integration decision-making described above, and on characteristics of “intendedly rational” decision makers more generally (Schmitt & Levine, 1977; Simon, 1976; Zedeck & Kafry, 1977), it was expected that managers would express conformance with a normatively preferable complex model of decision-making (Slovic & Lichtenstein, 1971), overestimating the extent to which they attended to some situational characteristics, underestimating the extent to which they relied on others, and in general, overestimating the extent to which they used the full range of information available to them. However, because managers primarily learn from experience (Levitt & March, 1988), it was expected that this difference between theories-in-use and espoused theories would be moderated by acquisition experience, such that managers with more acquisition experience would have developed a clearer set of integration-related values and greater understanding of the mental models created by them. Finally, it was anticipated that managers’ degree of insight into their theories-in-use in integration decision-making could be linked to acquisition outcomes.

The Decision Exercise

A set of decision scenarios was constructed in which the five factors suggested as influencing integration design decisions were embedded as decision criteria (i.e., the independent variables). The experimental design was based on a one-half fractional replicate of a full factorial design (Cochran & Cox, 1957) with each of the five decision criteria at high or low levels (e.g., particular acquisition situations were described as having high or low strategic needs, etc.), resulting in 16 decision scenarios per respondent. For each decision scenario, respondents were asked to indicate on a 7-point Likert scale the degree of integration (the dependent variable) they would recommend for that particular case (see appendix for a scenario example).
Statistical Analyses

The standard policy capturing approach was used to analyze the data. After estimation, the derived beta weights were transformed into relative weights to allow for a straightforward conversion into the percentage of total explained variance accounted for by each decision criterion (Hoffman, 1960). These objectively derived relative weights (reflecting theories-in-use) were then compared to subjective weights obtained by asking respondents to allocate 100 points among the various decision criteria according to their perceived relative importance to their integration design decisions (reflecting espoused theories).

Interviews

Respondents for whom completed questionnaires were received were asked to participate in a follow-up interview. Forty-nine out of the fifty-six participants in the study agreed to participate in the interview. The interview protocol was designed to elicit individual and organization-specific information related to the issues addressed in the decision-making instrument, but did not involve any discussion of the purpose or results of that exercise. The major categories of questions asked related to the nature of the respondents’ acquisition experiences, types of integration activities in which the respondent’s organization was typically involved, the key characteristics of acquisitions that determined the types of integration activities that would take place, the timing of integration activities, and the perceived causes of acquisition success and failure. The last category was intended to gain insight regarding the extent to which respondents attributed acquisition success or failure to issues relating to integration. Qualitative interview data were analyzed using content analysis and frequency counts of responses on questions for which a prespecified set of responses had been determined to be theoretically applicable (Ericsson & Simon, 1984; Holsti, 1968). For questions where prespecified responses had not been identified, responses were reviewed across participants to identify emergent patterns.

RESULTS

These managers were quite consistent in how they used situational information to make integration decisions ($R^2 = .67$), and as can be seen
from Table 1, they focused primarily on information relating to organizational needs. The data also indicate that these managers had little insight into their integration design decisions, with subjective relative weights differing significantly from objective relative weights for four out of five decision criteria.

Specifically, comparison of the managers’ objectively calculated relative weights to the subjective relative weights they reported they used indicates that these managers believed that they used a more equal weighting policy than the objective analysis of their decisions suggests. The calculated relative weights varied from 7.39 to 47.87 percent (Range = 40.48 percent), whereas subjective weights varied only from 12.64 to 28.13 percent (Range = 15.49 percent). This finding indicates that these managers believed their mental models and underlying value sets incorporated all five situational factors in relatively similar proportions, whereas in actuality their integration design decisions were largely explainable in terms of only two factors (i.e., organizational needs and strategic needs). Furthermore, the most important objective factors were subjectively viewed as less important than they were objectively indicated to be, while the least important objective factors were subjectively viewed as more important than they were objectively revealed to be. The correlation between calculated and subjective weights, across individuals, was only .53 (not significant).

The relationship between decision maker insight and acquisition experience was examined by correlating an insight index with acquisition experience across the 56 respondents. The insight index was obtained by correlating each respondent’s subjective and objective relative weights

<table>
<thead>
<tr>
<th>Decision Criteria</th>
<th>Mean Objective Relative Weights</th>
<th>Mean Subjective Relative Weights</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic needs</td>
<td>26.97</td>
<td>28.13</td>
<td>−0.33</td>
</tr>
<tr>
<td>Organizational needs</td>
<td>47.87</td>
<td>21.58</td>
<td>7.99***</td>
</tr>
<tr>
<td>Multiculturalism</td>
<td>7.39</td>
<td>15.79</td>
<td>−4.16***</td>
</tr>
<tr>
<td>Power differential</td>
<td>7.80</td>
<td>12.64</td>
<td>−2.33*</td>
</tr>
<tr>
<td>Compatibility</td>
<td>9.97</td>
<td>21.86</td>
<td>−5.69***</td>
</tr>
</tbody>
</table>

*p < .05.

**p < .01.

***p < .001.
across the five decision criteria. The Spearman rank-order correlation between the insight index and experience was $-0.13$, not significant and in the opposite direction to that hypothesized.

The qualitative data collected in the executive interviews confirmed these managers’ espoused policies, in that they indicated the critical importance to acquisition success of balancing strategic needs with a sensitivity toward cultural and compatibility issues. A statement related to compatibility made by one CEO is representative:

I think the first mistake you make up front is not getting clear on the relationship and the rules. If you get the rules up front and people understand that and they still want to sell to you, you have an environment to bring about necessary change. If they come in with misapprehensions and thinking it is going to be different from what it is … then they get all sorts of self-righteous feelings of having been abused. You have to get real clear, eyeball to eyeball, up front about how you are going to run it, and whether they are going to be willing to do it that way.

With regard to culture, a typical assessment made by a Vice President of Corporate Development was:

The culture factor is a significant and highly underestimated factor in bringing companies together …. We’re pretty sensitive to the culture of our company and the culture of the company we’re acquiring, and if we think there’s going to be a problem there, that would cause us to do things differently after the acquisition.

Respondents also recognized the importance of organizational needs, but their lack of insight into the extent of their actual reliance on this situational information may provide understanding of some of the acquisition failures these managers reported. Specifically, nearly 20 percent of these managers expressed a fear of compromising an acquisition’s value by blending it through integration, while simultaneously citing lack of control over acquired company actions due to insufficient integration-based coordination as a major source of post-acquisition problems. As one company president expressed:

I think it is very important not to begin making wholesale changes …. Although you want things to be more successful, you sure as hell don’t want to destroy it.

The CFO of another company put it this way:

I think there is a kind of fear we have that if we take something and homogenize it by integrating it, we are not diversifying our bets.
Yet, these same executives and others, in identifying causes of problems in acquisitions cited a lack of sufficient integration, as exemplified in the experience of this CEO:

We had an acquisition we left running autonomously and we ultimately discovered that the management wasn’t motivated and was moving in a different direction than the intent of the company as a whole. Ultimately we had to remove that management and integrate that business into the parent company, resulting in that business becoming profitable.

As illustrated by both the experimental and interview data, while the theories of integration to which these managers pledge allegiance should result in a balancing of diverse situational factors, the skewed theories actually guiding their integration decision behavior are reflected in a decision-making bias that appears to be impacting acquisition outcomes.

The results of this study suggest that managers are not aware of the extent to which their integration theories-in-use diverge from the theories of action they espouse. This lack of explicit reflection on the linkage between thought and action is blindingly apparent, as exemplified by the following quote:

(Integration) happens, but I don’t think with any particular strategy …. Once in a while we’ll integrate a lot more because maybe we’ll have a couple of things pop up, but no real strategic thinking going into it. (Executive VP Corporate Development)

This statement is illustrative of the general findings of this study regarding the degree of insight managers have into their integration decisions. That is, on average, the decision makers in this study had little insight into their decision models and underlying values regarding integration. This was evidenced by differences between what participants said was important and what their decisions indicated was important as reflected in the objectively determined weightings of the five decision criteria.

Greater levels of experience do not appear to ameliorate this lack of congruence. Experience was hypothesized to be positively related to decision maker insight into integration design decisions. The hypothesized positive relationship was based on acquisition research which suggests that experience is positively related to acquisition performance because experienced decision makers have a greater understanding of the key factors affecting integration success (Finkelstein, 1986; Kusewitt, 1985; Shanley, 1987). In this study, it was found that the direction of the relationship of experience to insight, although not statistically significant, was negative. The negative relationship may be interpreted as reflecting that with experience in this particular domain, decision processes become even more intuitive and inaccessible (Slovic, Fleissner, & Bauman, 1972), and
that competence in linking cognitions and actions is not developed through experience alone.

The lack of insight exhibited by these decision makers may result because decision makers recall having attended to each of the decision criteria, and thus they conclude that all influenced their judgments to some extent (Abelson & Levi, 1985). A possibility of greater potential significance for practicing managers is that they may report weighting available situational information based on “implicit theories” (Nisbett & Wilson, 1977) of what they think should be influential in making integration decisions. In other words, they may believe they are making integration decisions based on some comprehensive set of values that they have identified as normatively preferable, while their actual decision policy falls short of that ideal.

To the extent that acquisition failures can be explained by overzealous protection of the acquired firm’s autonomy and a concomitant lack of sufficient integration, awareness of these incongruities could be instrumental in the avoidance of problematic acquisition outcomes. To the extent that improved insight into integration decision-making can ameliorate problems of underintegration and a lack of focus on the requirements for realizing potential synergies, increased understanding of the links among managerial cognition, action, and performance in this domain should be of strong interest to management practitioners and researchers alike.

In attempting to understand the process whereby such incongruities result in less than optimal acquisition outcomes, our attention is drawn to two possible primary mechanisms. These include, first, communication-based difficulties with implementation of integration design decisions, and second, dissatisfaction with the “behavioral world” created by the decision maker’s theories-in-use (Argyris & Schön, 1974).

The centralized direction of organizational resources and member beliefs through strong executive leadership and communication is critical for the implementation of strategic change (Fubini et al., 2006; Tushman & Romanelli, 1985; Dutton & Duncan, 1987), particularly in such transformational events as acquisitions. As illustrated by the following statement, decision makers in acquiring organizations realize the importance of understanding and communicating why they have made certain decisions in order to generate the necessary acceptance and commitment to action among those who are charged with implementation.

You have to have a specific reason to do something, and you have to be able to articulate that to the organization . . . you better know what the hell you are doing . . . . Understand everyday, why did I do this, why did I do that, and if you can’t reinforce that, nobody’s going to believe it. (A Chairman of the Board)
This study’s findings regarding lack of insight suggest that such effective communication may be being short-circuited such that the values communicated as the bases for integration are in contradiction with the values actually underlying the integration design chosen for implementation. If this is the case, even the most well-intentioned transition managers will fail in their task.

A second process whereby incongruities between theories-in-use and espoused theories can result in unsatisfactory acquisition outcomes relates to what Argyris and Schön (1974) and Senge (1990) would characterize as the “problem of learning”. That is, the decision maker “must begin to make a connection between his own theory-in-use and those features of his behavioral world he most dislikes” (Argyris & Schön, 1974, p. 29). In order to change the unsatisfactory behavioral world of problematic integration and poor acquisition outcomes, it may be necessary to develop more effective theories-in-use by re-examining the values and ideologies underlying integration design decisions. However, until executives recognize the differences between their espoused theories and their existing theories-in-use, any connections that are made among underlying values, theories-in-use, and consequences will be in error.

The central practical implication of this study is that it suggests a method for making executives’ implicit integration decision policies explicit. Such cognitive feedback may be useful in helping managers recognize the divergence between their theories-in-use and their espoused theories. This awareness may help managers to purposefully focus on understanding how the values and ideologies they bring to the task of integration design may be impacting on acquisition outcomes. It may also allow managers to compare their own theories-in-use with other alternative theories that may be considered more suitable (Stahl & Zimmerer, 1984). In light of these improved insights, organizations could undertake a process of building preferred theories into formalized integration decision aids that could provide a useful communication and learning tool serving the organization’s future strategic needs.

As outlined in Table 2 below, by using policy capturing and complementing it with qualitative techniques, this study suggests one approach by which managers may be able to begin learning about the connections among cognitions, actions, and organizational performance in a very visible and important strategic domain. This chapter has illustrated that by bringing managers’ decision policies to light, they can become more effective caretakers of organizational health and stop sabotaging their own success. Lack of understanding of their own decision policies may be the greatest
saboteur to achieving this state. Further, researchers’ inability to recognize the decision policies in use will minimize their ability to analyze decision-making characteristics affecting processes and outcomes in a valuable way. That being said, we now turn our attention to the plusses and minuses of using a policy capturing methodology to understand strategic decisions.

### BENEFITS

One of the benefits of using policy capturing as a methodology for organizational research is that most management decisions involve the evaluation and integration of multiple decision criteria. Policy capturing is one method that allows management researchers to get a view, although indirect, of this process. The value of policy capturing in this instance is that it permits inference of a decision maker’s evaluation and integration of information by requesting assessments of a total situation rather than

---

**Table 2.** Take-Aways from this Chapter.

1. **When to use**
   
   This is a valuable methodological alternative:
   
   - if the intent is to understand decisions when multiple informational cues are available to decision makers
   - researchers want to learn what the influences of the different pieces of information are

2. **Research design when using**
   
   One can use this method to focus on:
   
   - the decisions of particular individuals, e.g., Bill Gates (idiographic research)
   - particular sets of individuals, e.g., CEOs (nomothetic research)

   Decision criteria (independent variables) – based on theory or field research
   The judgment or choice (dependent variable) – unidimensional

   Decision scenarios – vary in terms of scores on independent variables

3. **Avoiding problems**
   
   **Generalizability** – sample from the population to which you want to generalize, or use just for description rather than hypothesis testing
   
   **Construct validity** of “captured policies” – mathematical representation is just a paramorph of actual underlying decision processes
   
   **External validity** – the use of experimental designs employing hypothetical decision situations, and artificially orthogonal combinations of decision criteria – base decision situations on real-life examples
   
   **Sample** – choose those for whom decision situation is appropriate

4. **Improving research quality**
   
   Triangulate with qualitative data – interviews, archival data

---
requiring explicit evaluation of the various information elements. Thus, the focus of the analysis is on the decision of the individual rather than on the individual’s interpretation of his/her decision processes (Zedeck & Kafry, 1977). This is particularly valuable because managers are often unable (Simon, 1976), or unwilling (Pfeffer, 1981; Salancik & Pfeffer, 1978) to verbalize the reasoning behind particular decisions. Therefore decision-making research which uses direct questioning as a methodological approach may be problematic. Also, other policy capturing benefits are that it allows for systematic sampling of stimuli and control over confounding factors, precision of measurement, depending on the population of interest and sampling design used, and generalizability of results.

The value of policy capturing lies not only in allowing a decision maker’s implicit policy to be inferred and made explicit, but also in highlighting individual differences among decision makers (Slovic & Lichtenstein, 1971). The extent of individual differences among decision makers can be assessed by comparing the average level of decision consistency (average $R^2$) for individual decision makers to the level of decision-making consistency across the decisions for an entire group. Further, the ranges and variances of the weights associated with the various decision criteria are indicative of the degree of similarity or difference among individual decision makers’ policies (e.g., Rousseau & Anton, 1988; Stahl & Zimmerer, 1984). Clustering techniques can also be used to differentiate groups of decision makers by the within-group similarities and between-group differences in their policies (e.g., Bazerman, 1985; Naylor & Wherry, 1965; Wiggins, 1973).

Finally, the experimental designs typically used in policy capturing research allow for a measure of control and interpretational clarity frequently not available in other research methodologies. As such, policy capturing techniques are useful for shedding light on whether and how a particular set of variables appears to influence decisions in a specific domain, independent of the confounding influences of extraneous variables. This is particularly valuable for the study of phenomena that occur relatively infrequently, where systematic field variation is hard to find, and where the phenomenon of interest is fast-moving or confidential, causing access to key data to be problematic.

**LIMITATIONS**

Policy capturing is also subject to a number of limitations, which should be borne in mind when interpreting the findings of policy capturing research.
First, regarding decision makers’ “captured policies”, it is important to be conservative about the construct validity of this concept. That is, the mathematical representation of the judgment process is but a “paramorph”, or likeness, of decision makers’ actual underlying cognitive processes (Hoffman, 1960). More generally, a basic limitation of all decision-making research is that the mental process is not directly observable. As such, this process is “beyond the realm of legitimate scientific inquiry, except as it may be inferred from observable phenomena” (Hoffman, 1960, p. 116). However, to the extent that a functional relationship can be determined that accounts for consistencies in judgment in response to varying patterns of decision criteria (Hoffman, 1960), it is reasonable to discuss the decisions of individuals “as if” they were using a particular decision model. Thus, where clear theoretical models are available that predict observable decision outcomes, researchers using a policy capturing technique may reasonably speak of their findings as reflecting whether and how decision makers are using those theoretical models.

A second limitation of policy capturing research relates to the generalizability of findings. In policy capturing, a random sampling technique is not typically used to select the study’s participants. Rather, purposefully limited sampling is used to ensure that participants have the characteristics of the individuals whose decisions the researcher wishes to understand. As a result, in policy capturing, the results of regression analysis are typically used for description rather than for hypothesis testing, and the ability to generalize, or make inferences to a population beyond the sample is limited. Strictly speaking, the only population to which inferences may be made is the population of similar decisions the decision makers in the sample might make (Hays, 1981). In practice, however, the results also form a useful starting point for theorizing about how similar populations of decision makers might approach a similar set of decisions.

Inference based on regression analyses in policy capturing research is also problematic because the statistical reliability of “captured policies” may be jeopardized by low ratios of dependent variable observations to independent variables. The result is overfitting of the linear model and large sampling error (Hobson & Gibson, 1981; Tabachnik & Fidell, 1983).

Other characteristics of policy capturing that have negative implications for external validity relate to the use of experimental designs employing hypothetical decision situations, and artificially orthogonal combinations of decision criteria. Questions have been raised about the extent to which the evaluation of hypothetical decision situations corresponds in any meaningful manner to evaluations of similar actual situations. Despite Brown’s
(1972) empirical research demonstrating the lack of judgment differences under contrived and natural conditions, concerns remain about the effects of information presentation, specifically with regard to social context, volume, specificity, and format (Hobson & Gibson, 1981).

Furthermore, it can be argued that the computational and interpretational clarity associated with orthogonal designs comes at the expense of an accurate representation of the decision-making domain (Dudycha & Naylor, 1966; Hobson & Gibson, 1981). That is, there may be intercorrelations among decision criteria which exist in reality but that are not accurately represented in orthogonally designed policy capturing studies. For example, a study on selection might use age and work experience as two decision criteria. In reality, these two decision criteria are highly positively correlated, but in an orthogonally designed policy capturing study, in addition to scenarios depicting this natural positive intercorrelation, decision scenarios combining low age with high work experience would be also be included. In such cases as this, the external validity of the decision scenarios must be seriously questioned.

Another caveat to interpreting the meaning of captured policies is that it is important to bear in mind the implications for comparing across theories. Although the results of a study suggest that participants weighted certain of the decision criteria more heavily than others, and that they combined the decision criteria in particular ways in making decisions, we can only infer that these findings accurately describe the cognitive processes of the participants. As such, it can only be said that these findings appear to support one theory more than another, but a definitive comparison of the theories based on directly observable data cannot be made.

Finally, there is often no assessment as to whether each of the decision criteria was equally salient to the participants in the study. This is typical of policy capturing research that uses decision criteria specified by the researcher. Thus, although the manipulations on all the decision criteria were effective, it is possible that certain of the decision criteria were more salient to decision makers than others, or that the decision criteria were differentially salient to individual decision makers due to past experience or other individual-level characteristics. This factor should be taken into account when making comparisons across theories.

CONCLUSIONS

Despite the significant limitations of policy capturing, often they are outweighed by the benefits. The use of a policy capturing technique in a
research design involves trade-offs which must be weighed based on the primary goals of the research. Policy capturing brings unobservable decision-making processes into the realm of possibility for scientific examination, and provides a mechanism for evaluating the independent effects of specific informational cues on decisions in a particular domain. The findings of research using a policy capturing technique must be interpreted with caution due to methodological considerations impacting on validity and generalizability, and the basic indirect nature of observing the phenomenon of interest.

REFERENCES


Jim Conarty, President of Lone Star Chemical Corp., is really mad. Several months ago he charged some of his top managers with finding a way to cut operating costs at the company’s processing plants. These managers recommended to Jim that Lone Star acquire TexChem, a small company that had developed a very innovative and efficient production process. If Lone Star could get hold of and learn to use this technology, they could cut their operating costs by 20 percent. These managers made clear to Jim that transplanting this technology into Lone Star would require a lot of cooperation and close work between the managers and employees of the two companies.

Now that he’s signed the deal, Jim is getting very strong signals that TexChem is expecting to operate pretty independently. They are more concerned with developing new products and processes than in optimizing the use of TexChem’s existing process at Lone Star. This just doesn’t fit with Lone Star’s plans for this deal at all. Lone Star’s management has always discouraged differences of opinion about how things should be done at Lone

---

**APPENDIX**

<table>
<thead>
<tr>
<th>Acquirer</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Lone Star Chemical Corp.</td>
</tr>
<tr>
<td>Number of Employees</td>
<td>11,328</td>
</tr>
</tbody>
</table>
Star – it is part of their managerial belief system that such differences lead to inefficiency and a lack of control. At the same time, Jim’s afraid that in the case of TexChem, if he takes away too much of their independence during consolidation of the two companies, he may lose the talented people he needs to help get the new production process successfully implemented at Lone Star.
TESTING ORGANIZATIONAL ECONOMICS THEORIES OF VERTICAL INTEGRATION

Kaouthar Lajili, Marko Madunic and Joseph T. Mahoney

ABSTRACT

This article classifies empirical research on vertical integration under four approaches – value-added-to-sales, qualitative–quantitative, input–output, and microanalytic. The emphasis here is on the microanalytic approach which has accumulated the most systematic evidence to support its theoretical propositions. In particular, this article emphasizes theoretical and empirical contributions from organizational economics (especially transaction costs and agency theories) for both vertical integration and (vertical) contracting. Limitations and methodological challenges concerning the empirical testing of theories of vertical integration are addressed and suggestions for further research are provided.

INTRODUCTION

Why are some firms highly vertically integrated, while others specialize and outsource their remaining transactions in markets? A fundamental response
proposed by Coase (1937) maintains that the parties to an exchange take a comparative assessment and choose the governance structure (e.g., spot market, contract, hybrid, and firm) that reduces their transaction costs. Williamson (1971, 1975) emphasized that the Coasean transaction costs proposition required constructs that are operational. In particular, discrete structural forms need to be identified that have differential efficiencies, and the observable dimensions of transaction costs need to line up with various governance structures in a discriminating way (Williamson, 1991a).

This article focuses on the governance structure of the vertically integrated firm. The primary objective is to provide a framework for a systematic assessment of empirical literature in the fields of industrial organization, strategic management, and related fields that employ theories of vertical integration. Developing such a framework serves at least two purposes: first, it provides a useful cognitive map of the empirical research on vertical integration, and second, it facilitates further inquiry for theoretical and empirical advancement. The article is organized as follows.

The first section provides theoretical foundations for vertical integration, which are based primarily on transaction costs and agency perspectives. Empirical research is classified under four categories: (1) value-added-to-sales; (2) qualitative–quantitative hybrid; (3) input–output; and (4) a microanalytic approach. The second section focuses on the microanalytic approach. In particular, this section explores the testable implications of agency and transaction costs theories for explaining and predicting vertical integration. Empirical evidence from strategic management, marketing, and organizational economics perspectives are examined. Strong empirical evidence supports the conclusion that microanalytic empirical research yields systematic results for explaining and predicting vertical integration. The third section discusses limitations and methodological challenges concerning empirical testing of vertical integration. The final section provides a summary and suggestions for further research.

THEORETICAL FRAMEWORK FOR VERTICAL INTEGRATION AND CONTRACTING DECISIONS

The strands of the research literature (especially mathematical economic models) formally show the Coasean logic that in the absence of transaction costs, vertical contracting (e.g., exclusive dealing, resale price maintenance, and exclusive territories) can replicate the economic advantages of vertical
integration (Blair & Kaserman, 1983; Holmstrom & Tirole, 1989; Mahoney, 1992). Therefore, the formulation of vertical integration strategies (Harrigan, 1984) requires consideration of governance structures to implement business objectives (such as increasing revenues, decreasing costs, and reducing risks in ways that cannot be easily replicated by shareholders).

The generalizable thesis of the transaction costs research literature is that the particular governance structure chosen to implement the strategy of vertical integration primarily serves efficiency purposes (Williamson, 1991b). Williamson’s (1975) seminal research develops a well-grounded theoretical framework for explaining and predicting market failure. In short, contractual difficulties arise when opportunistic agents engage in frequent transactions in an environment of sufficient uncertainty and complexity that surpass bounded rationality capabilities (Simon, 1978). Furthermore, it is essential to underscore that environmental uncertainty and complexity, which can lead to incomplete contracting, allow for potential expropriation of economic quasi-rents only when relationship-specific investments surround an exchange (Klein, Crawford, & Alchian, 1978; Williamson, 1985).

The importance of relationship-specific assets in explaining and predicting vertical integration is supported by a large body of research literature including statistical testing (the primary focus of the current article) as well as formal modeling (e.g., Gibbons, 2005; Kleindorfer & Knieps, 1982; Riordan & Williamson, 1985) and case studies.¹ In contrast, research on vertical integration within the early industrial organization framework focused primarily on measurement techniques. The focus of the measurement literature was on relative comparison of industries with one another, or examination of firms and industries over time. Table 1 provides a list of empirical research on vertical integration including tests of transaction cost theory of vertical integration using (1) value-added-to-sales (e.g., Levy, 1985); (2) qualitative–quantitative hybrids (e.g., Armour & Teece, 1980); (3) input–output (e.g., MacDonald, 1985); and (4) microanalytic approaches (e.g., Masten, Meehan, & Snyder, 1991).

Transaction Costs and Agency Theory

A parsimonious framework that may explain and predict the choice of governance structure is developed here. The governance choice is influenced by frequency, uncertainty (demand and technological), and asset specificity (physical, human, and site) in transaction costs theory (Williamson, 1979).
<table>
<thead>
<tr>
<th>(1) Value-added to sales</th>
<th>(2) Qualitative-Quantitative Hybrid</th>
<th>(3) Input-Output</th>
<th>(4) Microanalytic (TCE, Measurement, Agency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crandall (1968)</td>
<td>Levy (1985)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lindstrom and Rozell (1993)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harrigan (1985b)</td>
<td>Muris et al. (1992)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frank and Henderson (1992)</td>
<td>MacDonald (1985)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harrison et al. (1990)</td>
<td>Maddigan and Zaima (1985)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The positive agency theory literature (Alchian & Demsetz, 1972; Eisenhardt, 1989) emphasizes the role of measurement uncertainty influencing governance choice. As different individuals organize activities into team production, monitoring of coordinated activities becomes a central problem. Asymmetric information (between principals and agents) due to team production leads to the so-called “nonseparability problem” (Alchian & Demsetz, 1972). If reward cannot be based on output, a manager will need to monitor behavior or effort (Barzel, 1982).

A second agency theory variable concerns knowledge of the transformation process or task programmability (Eisenhardt, 1985; Ouchi, 1979). Low task programmability reduces effectiveness of monitoring effort. The joining of transaction costs and agency theory yields frequency, asset specificity, demand uncertainty, technological uncertainty, task programmability, and non-separability as six key factors influencing governance choice (Mahoney, 1992). Although each of these variables has been operationalized, no single empirical study has considered all six variables simultaneously. The following section provides a microanalytical approach to develop propositions concerning vertical integration, which are the theoretical foundations to motivate implementing such an empirical study.

A MICROANALYTIC APPROACH TO VERTICAL INTEGRATION AND PROPOSITIONS

By selecting a particular governance structure, management aims to minimize the sum of production and transaction costs. This section advances 10 propositions based on organizational economics theories of vertical integration. Extant empirical evidence consistent with the outlined propositions is provided.

Microanalytic Approach: Propositions on Vertical Integration

Vertical integration can be viewed as substituting contractual or market exchanges with internal coordination of transactions. Specifically, such internal transactions are coordinated by an entrepreneur-coordinator who manages not by use of the price system but rather by fiat, which can substantially reduce the time and money that may otherwise be expended in the haggling between separate contractual parties. With this economic motivation in mind, it follows that vertical integration does not offer
advantage over a contract for a one-time exchange; however, as the frequency increases, the cost of vertical integration is more readily recovered. This economic logic suggests that increased frequency will increase the likelihood of vertical integration (Williamson, 1985).

**Proposition 1.** Vertical integration is a more likely governance choice when there is a high *frequency* of transacting.

Empirical evidence that supports this proposition can be found in Anderson and Schmittlein (1984), Heide and Miner (1992), and Klein (1989).

Transaction cost theory posits that contractual arrangements become more difficult to specify *ex ante* when uncertainty surrounding the exchange increases. Contracts designed under such conditions are necessarily incomplete and may require renegotiation in the face of unforeseen circumstances. Renegotiation poses a potentially hazardous threat for a contractual party who has limited exchange alternatives. Such an economic situation is known as small-numbers bargaining and because of the increased concern about contractual hold-up problems, there is anticipated to be an increase in the likelihood of vertical integration (Williamson, 1975). This dominant logic of vertical integration being a substitute for contracts when there is greater anticipation of contractual hazards leads to our second proposition.

**Proposition 2.** Vertical integration is a more likely governance choice when there are *small numbers* of potential trading partners.


Williamson (1996) identifies four basic types of asset specificity, and Masten et al. (1991) add a fifth type known as the temporal specificity. *Human asset specificity* involves uniquely related learning processes or teamwork. *Physical asset specificity* includes requirements for specialized machine tools and equipment. *Site specificity* occurs when unique locational advantages exist, as, for example, when a power plant is located near a coal mine to save on transportation costs. *Dedicated assets* are supplier’s general investments that would not have been realized but for the prospect of selling a significant portion of product to one buyer. *Temporal specificity* refers to assets that must be used in a particular time period. For example, even small delays in delivery of a certain production input can cause large economic losses (e.g., a newspaper company not integrated into press printing may incur (temporal)
hold-up problems). Vertical integration can assure requisite inputs in such situations. This economic logic leads to our third proposition.

**Proposition 3.** Vertical integration is a more likely governance choice when there is a high level of *asset specificity* (human, site, or physical capital and dedicated capital), which locks trading partners into a small-numbers trading situation that may make contracting hazardous due to potential haggling costs and “hold-up” problems.

Empirical evidence that supports this proposition can be found in:


(iv) Temporal or spatial proximity: Hubbard (2001), Masten et al. (1991), and Pirrong (1993).

Researchers have considered many types of uncertainty in the analysis of governance choice. Here we examine the effects of four types of uncertainty – demand (volume), technological, output measurement, and input measurement. Firms often face environmental uncertainty in the form of demand (volume) volatility (Walker & Weber, 1984). However, to the extent that volatile sales are anticipated, fluctuations in demand will not necessitate vertical integration, since a contingent claims vertical contract will suffice. Moreover, when asset specificity is low, competition attenuates opportunism, and hence demand uncertainty is inconsequential for the choice of governance structure. However, when asset specificity is high, an increase in volume uncertainty will have a direct positive influence on the likelihood of the governance choice of vertical integration (Williamson, 1979) due to increased contractual costs relative to hierarchical coordination. The
economic logic that vertical integration is more likely to substitute for contracts when uncertainty is high since contracts will be more incomplete and thereby pose greater contractual hazards leads to our fourth proposition.

**Proposition 4.** Vertical integration is a more likely governance choice when there is higher demand (volume) uncertainty, which makes contracting more hazardous (under conditions of asset specificity).


Increased technological uncertainty, which we turn to now, leads to different dynamics than demand uncertainty. The uncertain timing of the obsolescence of a technology can lead the firm not to choose a highly firm-specific technology, and hence vertical integration is less likely. From a real options perspective (Sanchez & Mahoney, 1996), the firm under uncertainty may not want to exercise its option to commit to vertical integration. As technological uncertainty is resolved, the sunk cost commitment to vertical integration may be made. This dominant logic leads to our fifth proposition.

**Proposition 5.** Vertical integration is a more likely governance choice when there is low uncertainty about the timing of the obsolescence of specific assets since this condition will allow greater investment in specific assets, which increases the likelihood of vertical integration.


If the uncertainty is due to the complexity of coordinating a technological system and transferring information (Teece, 1980), then vertical integration has typically been the predicted governance structure. The economic logic is the standard one that with increased complexity, contracts will be more incomplete and thereby pose greater contractual hazards (Grossman and Hart, 1986). In terms of empirical corroboration, Masten et al. (1991) find empirically that the strong association between human capital specificity and the increased likelihood of vertical integration is a consequence not so much of a decrease in the internal costs of organization, but rather is due to an increase in the cost of market exchange. Similarly, subsequent empirical evidence has found that the ease with which unstructured technical dialogue
is carried out between different departments of a semiconductor factory induces a need for hierarchically organized exchange (Monteverde, 1995). Both economic logic and empirical observation leads to our sixth proposition.

**Proposition 6.** Vertical integration is a more likely governance choice when there is increased complexity, which necessitates a higher degree of complex firm-specific language and routines.


**Agency Costs**

In terms of the “efficient boundaries problem” (Afuah, 2001; King, 1992; Ouchi, 1979), ease of effective monitoring of work behavior favors vertical integration (Anderson & Oliver, 1987; Ouchi, 1979). Eisenhardt (1985, 1989) considered four measures of task programmability (service, product, selling time, and training time) and finds task programmability a significant influence on governance choice. If environmental conditions are uncertain and consequently outcome uncertainty is high, then it is difficult to determine effort from observing output (Eisenhardt, 1985; Lassar & Kerr, 1996). To the extent that improved monitoring of input is effective (i.e., high task programmability), vertical integration is predicted. This agency theory logic leads to our seventh proposition.

**Proposition 7.** Vertical integration is a more likely governance choice when there is higher task programmability, which allows for effective monitoring of inputs.

Empirical evidence that supports this proposition can be found in Eisenhardt (1985) and Jones (1987).

In addition to environmental uncertainty, transactions (agency) costs may arise from behavioral uncertainty. A significant aspect of information asymmetry in organization is the problem of ascertaining and rewarding individual effort in team production (Jones, 1984). Outcome uncertainty may be due to free-riding behavior in team production – the so-called non-separability problem (Alchian & Demsetz, 1972). While the source of uncertainty is now behavioral rather than environmental, the contractual
problem is still the same: Output is not a sufficient statistic for inferring individual effort. Once again, vertical integration is predicted.

**Proposition 8.** Vertical integration is a more likely governance choice when there is a high non-separability problem, which thereby necessitates that inputs be monitored to determine individual productivity.

Empirical evidence that supports this proposition can be found in Anderson (1985), Anderson and Schmittlein (1984), John and Weitz (1988), and Poppo and Zenger (1998).

To the extent that observation of output is not satisfactory for completing a market transaction, the monitoring of inputs and vertical integration may be favored to minimize costs. Goldin (1986) notes that it is generally presumed that one can monitor output quality more cheaply in lower-quality goods than in high-quality goods. In the latter, one may want to screen workers and hire only those who will produce goods of uniformly high quality and then supervise only by input (i.e., hierarchy). Such was the case in the manufacturing of clothing at the turn of the century; high-quality coats, for example, were made by skilled tailors working on time (i.e., salary), while lower quality coats were made by piece rate via independent workers. Relatedly, vertical integration may be an adaptive response to a product differentiation strategy that is driven by changing customer demand or technology supply conditions. For example, Barry, Sonka, and Lajili (1992) note that product differentiation at the farm level (e.g., corn with high oil content, soybeans designed for specific international markets) may lead to different quality control and monitoring costs for which new vertical coordination organizational forms may evolve. This agency theory logic leads to our ninth proposition.

**Proposition 9.** Vertical integration is a more likely governance choice when there is a higher degree of difficulty in ascertaining quality of (a differentiated) product by inspection, which suggests that the monitoring of inputs is required.

Empirical evidence that supports this proposition can be found in Anderson (1985), Anderson and Coughlan (1987), Caves and Bradburd (1988), and Jacobides and Hitt (2005).

Finally, a major proposition of transaction costs theory (Williamson, 1985) is that vertical integration is most likely to be chosen when both uncertainty and asset specificity are high, since contractual hazards are likely to be the most severe. Thus, conditions of increasing uncertainty have a positive effect
on the likelihood of vertical integration conditional on the presence of asset specificity. This fundamental economic logic leads to our tenth proposition.

**Proposition 10.** Vertical integration is a more likely governance choice when there is an interaction of high uncertainty and high asset specificity. Empirical evidence that supports this proposition can be found in Anderson (1985), Coles and Hesterly (1998), Leiblein and Miller (2003), Leiblein, Reuer, and Dalsace (2002), Villalonga and McGahan (2005), and Walker and Weber (1987).

To be sure, there is substantial empirical evidence that corroborates a micro-analytical organizational economics approach for explaining and predicting vertical integration (Mahoney, 2005). We turn next to addressing some of the limitations and methodological challenges concerning the empirical testing of these theories of vertical integration as we move forward.

**LIMITATIONS AND METHODOLOGICAL CHALLENGES**

The first two sections of the current paper have focused on improving the model specification for explaining and predicting vertical integration from an economic efficiency perspective, and recent research shows improved model specifications are being adopted (e.g., Parmigiani, 2007). We also hasten to add here that it may prove fruitful to provide research designs that enable comparison of rival explanations in which alternative hypotheses are compared. For example, Spiller (1985) compares asset specificity and market-power explanations for vertical integration. Poppo and Zenger (1995, 1998) and Nickerson, Hamilton, and Wada (2001) also provide exemplar research designs for comparative examination of alternative theories.

Beyond econometric model specification, another important class of empirical problems involves econometric identification problems. Hamilton and Nickerson (2003) correctly note that a fundamental challenge in strategic management research is correcting for endogeneity. In the context of empirical work concerning governance choice for vertical coordination, strategic governance decisions are not made randomly, but rather are based on expectations of how these governance choices will influence future economic performance. Put more precisely, management’s
governance decisions are endogenous to their expected economic performance outcomes (Masten, 1993). In other words, many empirical papers attempt to answer the (implicit) question: “How does the economic performance of firms that made a particular governance choice compare to that of firms that made alternative governance choices?” However, from a governance-choice perspective, the correct social science question that needs to be answered is: “How does the economic performance of a firm that made a particular governance choice compare with how the same firm would have performed if it had adopted an alternative governance choice?

The endogeneity problem has substantive implications concerning our statistical analysis of these governance decisions. Statistical analysis that does not take into account management’s expectations of economic performance outcomes with respect to their governance decision can result in biased coefficient estimates. These biases result from key omitted variables that influence both governance choice and economic performance. Therefore, it is important that researchers in the strategy field utilize state-of-the-art econometric methodologies that account for omitted variables. Such econometric methods to correct for endogeneity when both strategic choice and economic performance are continuous include instrumental variable and two- and three-stage methods. In addition, econometric techniques to correct for endogeneity arising from discrete governance choices are growing in number as new econometric advances are made to correct for management’s self-selection of their (discrete) governance choice (Heckman, 1974, 1979; Lee, 1978, 1982). Exemplar research to guide current empirical work include Masten et al. (1991), Ohanian (1994), Poppo and Zenger (1998), Shaver (1998), González-Díaz, Arruñada, and Fernández (2000), and Saussier (2000) among others. We next consider further challenges in the discussion and conclusions section.

DISCUSSION AND CONCLUSIONS

Since the canonical problem in transaction costs theory concerns vertical integration, the final section focuses on areas for improvement here. Empirical research has provided strong support for central predictions of the transaction costs theory (Joskow, 1988a; Macher & Richman, 2006; Shelanski & Klein, 1995). Specifically, empirical findings generally corroborate the importance of various forms of relationship-specific
investments for explaining and predicting vertical integration. However, important issues, primarily associated with relationship-specific variables, invite further refinements. First, many empirical papers addressing the relationship between some form of asset specificity and vertical integration employ crude secondary data sources to approximate the underlying asset-specific investments. For example, researchers often utilize available measures such as advertising-to-sales ratios and R&D expenditures-to-sales ratios in the hopes of capturing different forms of asset specificity. Macher and Richman (2006) suggest, however, that when such crude proxies of asset specificity are found statistically significant, it is difficult to separate the effects of a particular variable from other confounding factors that may correlate with the specified explanatory variable. Moreover, multiple measures of explanatory variables are needed to improve construct validity of the analysis.

Second, additional difficulties in interpretation of results arise from employment of sub-optimal variables in a particular empirical setting. As Oxley (1997) suggests, many empirical studies relying on transaction cost rationale use firm-level characteristics to approximate for the transaction-level characteristics outlined by the theory. Oxley (1997) drawing from Williamson’s work (1985), emphasizes that the microanalytic attributes of transactions, and not firm attributes, influence governance choices and should be used in empirical work. Much of the research that examines governance modes in international expansions has been susceptible to the critique of being a “non-microanalytic” approach.

A third lingering issue concerning empirical research on vertical integration (and also other governance forms) relates to empirical treatment of relationship-specific explanatory variables. Routinely, asset specificity variables have received an exogenous specification within the logistic model intended to explain governance choice. Related to endogeneity problems discussed in greater detail in the previous section, firms’ decisions to invest in relationship-specific assets and to determine the amount invested are endogenous decisions (Masten & Saussier, 2002; Riordan & Williamson, 1985). Only a handful of recent papers such as Lyon (1995) and Saussier (2000) correct for this endogeneity problem.

**Directions for Future Research**

Internal costs of organization may play a significant role in integration decisions. Many empirical make-or-buy tests cannot distinguish if observed
governance forms are a consequence of market exchange hazards or are due to some systematic variation in the internal costs of organization (Masten et al., 1991). In this regard, Gibbons (2005) maintains that a more inclusive, and hence more appropriate, test of governance choice include a unified account of both the costs and the benefits of vertical integration.

Another direction for future research is to study vertical integration from the perspective of path dependencies and interdependencies (Mayer, 2006; Rothaermel, Hitt & Jobe, 2006; Schilling & Steensma, 2002). For example, Argyres and Liebeskind (1999) maintain that governance non-separability – i.e., interdependencies between related governance choices – can play an important role. Indeed, the formal and informal constraints embedded within the firm’s existing set of contractual commitments can influence subsequent governance decisions. Empirical research that joins institutional theory in organization theory with institutional economics appears promising. Relatedly, Leiblein (2003) notes that to the extent that resources and capabilities might be operationalized as clusters of transactions, approaches that consider multiple transactions through some forms of interdependence may facilitate the integration of transaction cost theory and dynamic resource-based theory.

Finally, a fruitful direction for research concerns assessing how changes in information technology influence governance choice. Indeed, recent changes in coordination technologies can substantially impact transaction costs. For example, complex products previously requiring intensive coordination through in-house development and production are now being handled by loosely coupled processes for which coordination across many participating firms is now transaction cost efficient (Lajili & Mahoney, 2006; Sanchez & Mahoney, 1996).

Researchers in the strategic management field have already begun to address some of limitations emphasized here. We conclude with the anticipation that with better theory development and better econometric techniques, the next generation of researchers in the strategy field will take the existing research, and will do better.

NOTES


2. Transaction costs include the \textit{ex ante} costs of (1) search and information costs; (2) drafting, bargaining, and decision costs; and (3) costs of safeguarding an agreement. Ex-post costs include: (1) costs of measuring input and output; (2) monitoring and enforcement costs; (3) adaptation and haggling costs; (4) economic bonding costs; (5) mal-adaptation costs; and (6) the residual economic loss due to shirking and cheating.

REFERENCES


Testing Organizational Economics Theories of Vertical Integration


AN ONTOLOGICAL FOUNDATION FOR STRATEGIC MANAGEMENT RESEARCH: THE ROLE OF NARRATIVE

Dominik Heil and Louise Whittaker

ABSTRACT

The double challenge of strategic management research is to come up on the one hand with a truthful account of issues relating to the organization itself and on the other hand doing so in a way that is relevant for meeting the challenges of strategic management. To achieve this need for increased rigour and relevance we follow Heidegger’s thinking and develop an ontological ascertainment of the organization as being a work. A work in this sense is the kind of entity that is fundamentally characterized by setting up a world for people. We then argue that because an organization is a work, an appropriate way of giving an account of an organization is in the form of strategic narratives: narratives that give a truthful account of the world of an organization as it is and as it could be. Since narratives play a fundamental role in human existence and are a powerful means to shaping peoples thinking and actions, we argue that strategic narrative research and development meet the double challenge of strategic management research and illustrate the application with an account of strategic narrative research and development in an organization.
INTRODUCTION

For many managers, academics who write research papers ‘say nothing in these articles and they say it in a pretentious way’ (Ives, 1992). In other words, the work that is done in management research is often perceived to lack relevance in the eyes of practitioners, who are the very people who should benefit from this kind of research, because strategic management is a discipline based in practice. We suggest that the cause of this lack of relevance is not a function of too much but of too little rigour in strategic management research. The rigour that we regard as missing however is not the rigour of applying sophisticated research methodologies correctly from an epistemological point of view, but rather the rigour in confronting a logically more primordial – though often overlooked – question about the very nature of the organization as the entity that is ultimately being researched. We contend that appropriate research methodology would first and foremost need to be in harmony with the nature of the entity that is to be researched.

To ask the question ‘what is the very nature of an entity?’ is to ask the ontological question. While we find aspects of the ontological question in strategy and other aspects of business thought (i.e. Nonaka, Peltokorpi, & Tomae, 2005; Sullivan, 1998), we find that a comprehensive ascertainment of the very nature of the entity called ‘organization’ which is grounded in the tradition of ontological thought is missing.

The purpose of this chapter is to contribute to both the rigour and relevance of strategic management research by developing an ontological understanding of the very nature of the organization along the lines of arguably the most prominent contributor to the field of ontology: Heidegger. Within Heidegger’s thinking an organization is ascertained as being in its very nature a ‘work’ (Heil, 2003). The word ‘work’ is being used here not in the sense of labour but in the same way as one talks about a ‘work of art’, the Latin word ‘opus’ or the French word ‘œuvre’. A work in this sense is the kind of entity that is fundamentally characterized by setting up a world for people, a notion we will explain in more detail later in this chapter. Ascertaining organizations as works has significant implications on the one side how a truthful account can be made about this kind of entity and on the other how this account can be made in a way that is relevant to the challenge of managing this type of entity. After developing and explaining the notion of the organization as ontologically being a work we proceed to suggest the research and development of strategic narratives as a research methodology that is in harmony with this very nature and meets the challenges of giving a truthful, rigorous and relevant account of companies (and other
organizational institutions), which we complement with an account of the actual application of this methodology in an existing organization.

**RIGOUR AND RELEVANCE IN STRATEGIC MANAGEMENT RESEARCH**

‘Rigour’ is an overarching term that we use to justify the knowledge produced by our research. The details of this justification will typically depend on the discipline and philosophical grounding of the research, and will appear as more or less problematic in various fields. In the positivistic tradition, criteria of validity and reliability are used ‘to measure the extent to which our theories and instruments correspond to objective reality’ (Sandberg, 2005, p. 43). In interpretive approaches, various criteria have been proposed by authors such as Sandberg (2005), Whittaker (2004), Prasad (2001), Klein and Myers (1999) and Madison (1990). While these differ in detail, and are nowhere near approaching the universally recognized status that reliability and validity hold in positivist research, they do agree on a basic premise that the truth claims of a piece of research must be consistent with the assumptions underlying the research (Sandberg, 2005).

Relevance on the other hand has to do with helping managers to make better decisions. Managers do not have any particular interest in how we came to know about what they have done in the past, and often the methodological improvements that we make in the interests of increased rigour are of no interest to them, but actually make our research opaque and therefore, less relevant. Thus, as the field of strategic management undergoes a rapid transformation in methodological rigour the challenge of maintaining relevance will become more obvious and acute.

This is not a challenge that should be trivialized. Currently there are a number of significant critiques of management research and theory. Some critique suggests that management research lacks relevance and is even destructive, particularly when viewed along ethical dimension (Goshal, 2005). We would contend however, that the shortcomings in strategic management research relevance should not be exacerbated by improved rigour, but that a rigorous consideration of what strategic management research is about in the first instance should significantly improve relevance.

This is, of course, by no means an unexplored issue in the strategic management literature. For example, in a series of articles in the *Strategic Management Journal*, Powell (2001, 2002, 2003) points out that the
philosophical foundations of strategy research are assumed, unexplored and more seriously, ‘not true or real or acceptable – either scientifically or practically – under any positivist or anti-positivist epistemology’. Powell (2001) suggests that the way to provide at least ‘a consistent intellectual foundation’ is to adopt a pragmatist perspective and thus ‘resist the vanity that we are gaining on objective truth or reality’ (p. 886). The pragmatist approach, Powell (2003) suggests, allows us to adopt a broadly construed empiricism, without the pernicious consequences of the interjection (explicitly or otherwise) of ontological beliefs.

Pragmatism of course has its own ontological belief – that we cannot deal with ontology, that we should ‘stand clear’ of ontology – because neither positivism nor anti-positivism, or any other broadly empiricist approach can prove its own ontology. In this chapter, however, we want to suggest that an alternative to pragmatism – which avoids ontology altogether, is phenomenology – which tackles the ontological question head on.¹ Phenomenology (specifically Heideggerian hermeneutic phenomenology) is a radical ontological questioning of things that usually show themselves inconspicuously, that is of the phenomena of everyday objects, practices and discourse. By questioning the very basis, or underlying assumptions of these things we can ‘squeeze them into focus. Phenomenology is only needed because some matters … are hidden. Hidden not because we have not yet discovered them or have simply forgotten them, but because they are too close and familiar for us to notice or are buried under traditional concepts and doctrines’ (Inwood, 1999, p. 160). In order to understand specific matters (such as strategy), we need to question their very basis, or being. It is not enough to seek only to observe collections of empirical things – the ontic view – we need to seek their meaning as phenomena – the ontological view (Introna & Whitley, 1997).

Thus, in adopting a phenomenological approach to strategic management research, we reframe the question of rigour and relevance from the epistemological question to the ontological one: how do we understand that with which strategic management is concerned? What is strategic management about, and how can we know about that?

As soon as we do this, however, we encounter another layer of epistemology, or what Giddens (1984) refers to as the double hermeneutic of the social sciences, in which we try to understand how it is that people understand things. What we need to know about in strategic management research is ‘how managers make major decisions’, and even before they can decide, how they know about what it is that they are going to decide about. This of course, is why strategic management theory provides a plethora of frameworks for managers to use, all of which are ways of framing
knowledge. In as much, then, as strategic management practice is epistemological, it (or rather the practicing manager) necessarily makes ontological assumptions about that which is known or decided about. To make a decision, managers must make some assumptions about what they are deciding about. These assumptions necessarily underpin any epistemological process, and therefore profoundly affect the nature of that process (how managers make decisions) and any research about it (how we know how they make decisions). This double hermeneutic with underpinning ontological assumptions is illustrated in Fig. 1.

In order to provide both improved rigour and practice then, we need to enquire into the ontology of the object of strategic management practice. That is, we need to consider the very nature of the entity that strategic management is dealing with. As strategic management can be broadly defined as the management of an organization itself (Mintzberg, Ahlstrand, & Lampel, 1998), we contend that this entity is the organization itself. Therefore, the fundamental ontological question that needs to be posed first and foremost for research in this area is the ontological question of the very nature of the organization as an entity.

There is certainly existing literature on strategic management (Nonaka et al., 2005), and related fields such as international business (Sullivan, 1998) that discusses the ontological and epistemology assumptions of these domains. Thus far, however, the literature has not directly addressed and responded to the ontological question with a holistic understanding of the very nature of the organization in a comprehensive manner that takes the richness of the phenomenological tradition of ontological thought into account. That is not to say that there is not a considerable body of organization theory that provides a number of perspectives on how we can think about organizations, such as social constructionism (Fairclough, 1995; van Dijk, 1997), structuration theory (Giddens, 1984; Pozzebon, 2004), discourse

![Fig. 1. Epistemology and Ontology in Strategic Management Research.](image-url)
theory (Grant, Keenoy, & Oswick, 1998) and metaphorical analysis (Morgan, 1986). We would contend, however, that none of these perspectives directly asks the question, ‘what is a organization?’ as an ontological question.2

It is our aim, therefore, to suggest a unified understanding of what an organization actually is. The ontological question of the very nature of the entity to be researched should logically be the first question that is to be asked with regards on the appropriateness of any type of research methodology in a field of inquiry. In other words, instead of focusing the question of rigour at the top layer of our diagram in Fig. 1, we start at the bottom, with the primordial question: ‘what is an organization?’

THE ONTOLOGICAL ASCERTAINMENT OF ORGANIZATIONS AS WORKS

Heidegger3 from an ontological perspective distinguishes four types of entities: mere physical objects (such as stones); non-human organisms (such as plants and animals); human beings; and works. In investigating the fundamental characteristics of these types of entities as described by Heidegger (1984, 1992, 1994) we conclude that an organization is not of the same very nature as a mere physical object, a plant, an animal or a human being. The strategic management literature does frequently refer to companies as if they were of these types, for example as a machine (a physical object), as a growing thing (organism) or even as a legal person (a sort of abstracted human being). However these are all metaphorical accounts, and therefore by definition, not what an organization actually is. We might also suggest that an organization, even if it is not one of these things, can have all those types of entities within itself. This however is different from suggesting that an organization is actually one or a number of these entities. For example, one might suggest that the organization is a system (Beer, 1980; Kast & Rosenzweig, 1973; Katz & Kahn, 1978), a population (Hannan & Freeman, 1977), an ecosystem (Astley, 1984) or a self-organizing organism (Luhmann, 1995; Ulrich & Probst, 1984; Von Grogh & Roos, 1995). However there are at least two distinct differences between these collectives and the organization that suggest that organizations are not simply collectives of these types. Firstly, an organization is usually directed in a way that a system, population, ecosystem or organism is not, that is, with intent. Secondly, within the organization there are human beings who are aware of their role and their
existence within the organization, and who, as part of their very nature, interpret themselves within the organization (Kogut & Zander, 1996).

An organization, crucially, is real in the human experience, over and above its physical manifestation. Indeed, in the post-industrial era many companies have almost no tangibly physical manifestation. There are other entities that we encounter that do not, or do not primarily exist as a physical manifestation, and yet which are real in the human experience. Examples would be symphonies, Newton’s or Einstein’s laws of physics, nation states or poetry. There are also entities that have a very tangible physical manifestation, but whose significance in the human experience far exceeds only the physical structure, such as the Statue of Liberty, or the Eiffel Tower. These, Heidegger (1994) calls ‘works’.

Heidegger develops his notion of a ‘work’ in his seminal article ‘The Origin of the Work of Art’ first and foremost by way of the example of a particular type of work namely the work of art. He specifically mentions that there are other types of works such as poetry, music, art, religion, philosophy, architecture and states. We suggest that a organization too is a case of this type of entity: a organization in its very nature is a work (Heil, 2003). To explain the notion of the entity called ‘a work’, Polt uses the Vietnam Veterans’ Memorial in Washington, which was designed by Maya Lin and is usually referred to as ‘the Wall’ (Polt, 1999, pp. 135–136). The memorial is a simple V-shaped trench made from a series of black stone sheets inscribed with the names of all the American soldiers who lost their lives in the war. It has become a sacred site both in and beyond the United States. What makes it a work cannot be found in the material from which it is made nor in its beauty or aesthetic appeal, but in its effect of creating a world. The Wall makes the Vietnam War present, and establishes and preserves this event as a fundamental dimension of American identity regardless of the political conviction of the visitor to this work. Such revelations belong to every work. According to Heidegger (1994) to be a work means to be the kind of entity that sets up a ‘world’. While they all have physical properties, works cannot be fully understood via an analysis of these properties. They can only be understood appropriately as works by attending to the world they set up (Heidegger, 1994). In other words one can only understand a work by being in the world that is set up by the work and cannot be appropriately understood by a detached observer. World here does not refer to all objects in the external environment or universe as it would be in the Cartesian tradition, but is used in a similar way as that which for example allows for ‘the world of a mathematician’ and then means ‘the realm of possible mathematical objects’ (Heidegger, 1984) or the
‘corporate world’, which refers to the realm of possible objects and issues concerning or relating to the corporation. However, more generally, it is a world which permits the possibility for ‘a world’ in the sense of an understood external environment to be there in the first instance. World means the always already familiar horizon upon which everyday human existence moves with absolute confidence and within which humans make sense out of both their environment and themselves. World is the ultimate reference which cannot be explained by any other reference and can thus not be determined via explanation. In Heideggerian terms, world is the referential whole in which practices and entities have meaning at all. World allows for the possibility of explanation itself in the first instance. The references within the whole sustain meaning in the context of the particular world, because they relate to each other, by showing up as something within the world. For example, a hammer can be seen to be a hammer because it is used to drive a nail, for the purposes of carpentry, for the purposes of creating chairs on which we sit, and tables on which we eat. Without the practice of sitting at table, what would the hammer be?

What a work does is to set up a world and thereby making it possible for references to show up for humans, as they are experienced by humans. For example, a rose in bloom is experienced by many humans as a thing of beauty, but by a farmer (in the commercial flower industry) as a source of income or by a chemist (in the world of perfumery) as a source of a chemical compound. For the layman a hammer is a hammer, for the carpenter each hammer is of a particular kind. Importantly, for humans, everything shows up as something, depending on the particular world in which one is. And what things show up as is what Heidegger (1994) calls actuality, or truth. Thus, as a work sets up and produces a world it is also that which ‘truth establishes itself in’. A work can be said to ‘work when it sets up a world’.

World worlds and is being more fully than the tangible and perceptible realm in which we believe ourselves to be at home. World is never an object that stands before us and can be looked at. World is the ever non-objective to which we are submitted as long as we are [human beings]. (Heidegger, 1994, p. 30)

World is what gives reference and allows for something to show up at all. Thus, an organization (as a work) sets up a world and by doing so provides a regime of truth and realm of possibilities for action. Whereas the business world is the totality of references that are the general conditions of possibility for us to be business people (i.e. economic transactions, markets, buyers, sellers, suppliers, banks, money, shops, marketing, etc.), in any particular organization there are particular conditions of possibility which
apply. A corporation in which customers show up as customers, and furthermore as important, for example, is probably in a competitive business, rather than a monopolistic enterprise.

Because, unlike all other entities, human beings are always in a world, their way of being has been called ‘being-in-the-world’ (Heidegger, 1984). Every action and decision of human beings is grounded in the world in which it happens. Crucially, ‘the world should not be construed as ‘what-is-not-[human]’ (Gordon, 2006), since ‘worldhood’ is a constitutive part of what it is to be human. We are human because we are in-the-world.

By setting up a world, organizations – like all other works – set up that within which human beings can meet each other as humans. For humans to encounter each other they need to share a world. When this fails, we say about two people or communities of humans that they are ‘worlds apart’. Humans need works to become communities (which are different from collectives of organisms). Humans who do not encounter each other within a world set up by works would, as Schwan points out, not encounter each other as human beings at all. Humans only meet each other as humans in a world (Schwan, 1989, p. 19). Works create community, a sense of shared belonging, shared meaning and a shared understanding of what it is to be a human being. The work creates the possibility for a common history. Humans can only become fellow human beings when they meet each other in a shared world. Only then humans can understand each other, collaborate and work with each other. It is works that sets up the world in which humans can and do form relationships with each other as humans. It is works that allows us to be human, both as individuals and within a community (Schwan, 1989, p. 19). Thus by setting up a world, an organization stands in the closest relation to the very nature of human beings as being-in-the-world.

COMpanies AND STRATEGIC MANAGEMENT

Considering the ontological status of an organization as a work is not merely an interesting analytical exercise. We suggested that we need to understand the nature of the organization because this is the primordial question, the first layer in understanding strategic management as illustrated in Fig. 1. However, our ontological understanding of the organization has a significant implication not only for research in strategic management, which is the top layer of Fig. 1, but for strategic management itself. Understanding the organization as a work, we need to ask how such an entity is appropriately understood and handled. The appropriate relation to the work so as to understand it
according to its very nature according to Heidegger is called ‘attendance’. Attendance in the Heideggerian sense is understanding the work from being within the world that the work sets up. To attend to the work is a way of knowing by participating in rather than just observing the work in a detached way. Attending to the work in this case means to be involved in the world that is set up by the work. The proper way of attending to and being involved in the work is solely given by the world that is set up by the work itself.

Attending to the work means standing within the openness of entities that happens in the work. This ‘standing-within’ or attendance, however, is a knowing. Yet knowing does not consist in mere information and notions about something. Those who truly know entities know what they will to do in the midst of them. (…) The attendance to the work, as knowing, is a sober standing-within the awesomeness of the truth that is happening in the work. (Heidegger, 1994, pp. 54–55)

In fact, if we understand the organization as a work, then the strategic manager, who must make the strategic decision about the organization (which is always a work), is always in-the-world that the work sets up. The manager cannot appropriately understand (think about, decide about) the organization from some vantage point outside of this world at all. The second and third layers in fact do not exist separately and the ontology of the organization means that we must collapse the distinction between knowing about the organization and the organization itself, as shown in Fig. 2. Knowing requires being in the world of the organization. A decision detached from the organization is by definition a decision that does not know what it is deciding about. Therefore, strategic management cannot be deciding about the organization, where ‘deciding’ and ‘the organization’ exist on two different levels (as is Fig. 1). Competent strategic management can only ever be deciding in the world of the organization.

This collapsing of the epistemological/ontological dualism is not a move which then denies the necessity for an ontology of the organization itself. It is rather the same phenomenological collapsing of the subject/object
dualism that underpins Heidegger’s (1953) ontology, and which makes the ontological consideration of how it is that we can think and decide about something (or strategically manage it) in the first place the primordial question. This same ontological enquiry reveals the world as that which we are always-already-in, and organizations as works that always set up particular worlds that they (companies and everyone in them) are always-already-in. We cannot think outside of our being-in-the-world, and we cannot think or decide outside of action.

Heidegger (1949) reflects on action in his Letter on Humanism

We are still far from pondering the very nature of action decisively enough. We know action only as causing an effect. Its actuality is valued according to its utility. But the very nature of action is accomplishment. To accomplish means: to unfold something into the fullness of its very nature, to lead it forth into this fullness – producere. Accomplishable is therefore really only that, which already is.

In as much as strategic management, ‘making major decisions about organizations’, is both thinking and action, we suggest that it is always an act of accomplishment. It is never something which exists independently and is then applied to an organization.

Strategic management must therefore be the type of action that is in correspondence with the organization as a work and the world that it sets up. Following Heidegger’s thought on thinking and action and the ascertainment of the organization as a work, we suggest that strategic management accomplishes the relation of the organization and the world the organization is setting up. It will also set up an understanding of the very nature of the human beings who are in this world. It does not make or cause that relation. Strategic management brings this relation to the organization and the world it sets up solely as something handed over from the organization and its world as it is and as it could or should be. Such offering consists in the fact that, in thinking, the organization as a work and the world it sets up comes to language – is articulated.

Language, importantly, is not mere chatter. Rather, ‘[l]anguage is the house of Being. In its home the human being dwells’ (Heidegger, 1949, p. 5). Those who are involved in strategic management and those who create with words (thinking, deciding and speaking) in an organization are the guardians of this home in an organization as it is and as it could or should be. Their guardianship is the accomplishment of the manifestation of the organization as a work and the world it sets up insofar as they articulate and guard them in language. According to Heidegger, such action is presumably among the simplest and at the same time among the highest
(Heidegger, 1949). Strategic management in the world of the organization lets itself be claimed by the organization and the world it sets up so that it says the truth of this work and its world as it is and as it could be. Therefore saying the truth of the work, or giving an account of the organization is the essential task of strategic management.

An appropriate account of an entity must be derived from the very nature of the entity that is to be accounted for. It might be perfectly sufficient to describe a granite cube in terms of its material and the quantitative measure of its spatial dimensions. In contrast, to describe a human being purely in terms of material composition and spatial dimensions tells us little or nothing about that which is critically relevant in describing a human being as a human being.

As companies have been ascertained as works an appropriate account would be derived from its essential nature as setting up a world. It would be impoverished to give an account of a work merely within the quantitative language such as market value and other quantitative measures (Heidegger, 1994). It is not that this language is incorrect, rather than it does not say that which is essential to the work as a work.

An appropriate account of a work would essentially need to account for the world the work sets up. We suggest that an appropriate account for a world would itself set up a world and consequently be a work itself. In following Heidegger’s description of all works being essentially poetic, we suggest that an appropriate account of an organization would itself be poetic in a way that this account sets up the kind of world that the organization sets up. In its purest form this would be a poem, though not necessarily in the sense of poems in the form of rhythm and rhyme but in the sense of using language to set up a world as it is used in all of literature.

For this reason we conclude that any appropriate account of an organization as a work would come in the form of a narrative. Narratives have the ability to capture much of the totality of the referential whole of the interpretations which make up a world. Strategic management research is therefore about understanding the nature of organizations and the worlds they set up and about appropriately articulating a narrative that reflects that world.

CREATING AN ACCOUNT OF AN ORGANIZATION THAT IS RELEVANT FOR THE CHALLENGE OF MANAGING AN ORGANIZATION

While so far we have argued that narratives are an appropriate and truthful way of giving an account of an organization, to meet the double challenge
of strategic management research it is also critical to understand the role of narratives in the world of the organization itself. Boland and Tenkasi (1995), following Bruner, point out that narrative is not an idle luxury of teatime chatting, but a fundamental cognitive process through which communities of practice are constructed and maintained. Narratives are distinct from logical arguments, validated by criteria of interest and plausibility, rather than only logic and consistency. Rather than relying on demonstrable proof, they show how events and things might fit, given a particular cultural situation. Polkinghorne states that ‘narrative is a form of ‘meaning making’ … Narrative recognizes the meaningfulness of individual experiences by noting how they function as parts of the whole. Its particular subject matter is human actions and events that affect human being, which it configures into wholes according to the roles these actions and events play in bridging about a conclusion … The narrative scheme serves as a lens through which the apparently independent and disconnected elements of existence are seen as related parts of a whole’ (Polkinghorne, 1988, p. 36). For stakeholders inside and outside the organization then, the use of language as narrative may be used to tell a story which surfaces the implicit assumptions, or background of meaning on which actions are taken (Boland & Tenkasi, 1995), as well as the events themselves (Goldstein, 1992). This process is one of collaboration and social construction, since ‘shared narratives are obviously communal and thereby collaborative’ (Brown, 1998, p. 225). Narratives therefore serve not only informational functions, but also help participants develop situated skills, and establish identity within the broader community in and around a organization. Ultimately, ‘collective wisdom depends upon communal narratives’ (Blackler, 1995, p. 1037). As Boland and Tenkasi express it, the unique knowledge of a community ‘develops by refining its vocabulary, its theories and values and its accepted logics through language and action within the community of knowing’ (Boland & Tenkasi, 1995, p. 355). Studies such as Orr’s (1990) support the claim that the main source of knowledge in companies is narrative. Thus the suggestion that knowing is narrated in communities of practice refers not to a process of knowledge transfer, but to the very process of knowing. As Boje notes, ‘storytelling is the preferred sense making currency of human relationships among internal and external stakeholders’ (Boje, 1985). In other words, people in and around organizations tell stories, thereby creating meaning for themselves, all the time, anyway. In phenomenological terms, the work that is the organization is created through narrative – organizations essentially are told, and tell a story.
The task of strategic management then is to appropriately understand and alter the story that an organization is and tells. Strategic management is itself at its most fundamental level a form of storytelling (Barry & Elmes, 1997; Czarniawska-Joerges, 1996). Given the ascertainment of the organization as a work which in each case sets up a world, this is also in line with Heidegger’s thinking about narratives and their ability to alter the world we share. With reference to a particular type of narrative, namely the Greek tragedy he states:

The same holds for the linguistic work. In the tragedy nothing is staged or displayed theatrically, but the battle of the new gods against the old is being fought. The linguistic work, originating in the speech of the people, does not refer to this battle; it transforms the people’s saying so that now every living word fights the battle and puts up for decision what is holy and what unholy, what great and what small, what brave and what cowardly, what lofty and what flighty, what master and what slave (see Heraclitus, Fragment 52). (Heidegger, 1994, p. 29)

Narratives have the capacity to alter the world we find ourselves in, in the most fundamental ways. Consequently, the challenge of strategic management is essentially to tell a powerful story and brings all aspects of the organization in line with this new story. Such stories must be true (truthful), rather than alienating, and yet they must also have the capacity to convey strategic intent – that which is not yet true but can and should be true in the future. According to Gardner, a story that succeeds is fulfilling this challenge ‘builds on the most credible of past syntheses, revisits them in the light of present concerns, leaves open the space for future events, and allows individual contributions by the person in the group’ (Gardner, 1995).

As Shklovsky (Barry & Elmes, 1997) points out, stories work when they have both credibility (or believability) and defamiliarization (or novelty).

Given the ascertainment of the organization as a work we conclude that narratives are both an appropriate account of an organization and appropriate and relevant in helping to meet the challenges of strategic management. Narrative research and story development and storytelling in its various forms therefore meets the double challenge as set out at the beginning of this chapter.

**DEVELOPING A STRATEGIC NARRATIVE**

In taking up this challenge, we have created and applied a process of developing a strategic narrative for a major telecommunications organization in South Africa, having been briefed on the strategy initially by a senior
executive and later also the Chief Executive of that organization. In doing so we are engaging in an action research programme (rather than research alone). At the heart of our programme is an attempt to apply the understanding of the organization as a work to the action of strategic management. In the intervention at the telecommunications organization, our effort was to create an appropriate strategic narrative, where an appropriate narrative would be one which draws both on the fundamental nature of the organization as a work and the world that people within the organization find themselves in.

The organization where we have conducted the work is a recently privatized and newly listed (in 2001) telecommunications monopoly in South Africa. With the privatization and listing of the organization, a strong focus on shareholder value and performance enhancement has led to exceptional share price performance. Extensive retrenchment exercises have reduced the size of the organization from 40,000 to around 25,000 employees in 5 years. The customer and employee perceptions of the organization, as captured in various surveys, were less positive than the shareholders. The state-sanctioned monopoly of the organization was about to be ended with the introduction of a second network operator.

To ‘get at’ the organization and its world we followed Prasad’s (2001) hermeneutics as a methodology for understanding institutions, which is based on the notion that a organization itself can be read as a text, an approach that is both in line with and draws explicitly on Heidegger. To stimulate members of the organization to share this text with us we used questions with regards to those fundamental characteristics that articulate the dimensions for developing an account of the world of an organization. These are the questions that should surface that which is always-already known within the organization, but to some degree unsaid. In these questions, the very nature of the organization as a work, and the world that it sets up can be articulated. Consequently the types of questions that are to be asked are those that interrogate that world. Deriving these questions requires us to consider what gives us ‘what is’ in the world of an organization. Heidegger gives us a few hints in his seminal essay ‘The Origin of the Work of Art’ (Heidegger, 1994) about the key characteristics or dimensions of a world. The two most important passages of Heidegger’s essay for our formulation of the questions were:

As a world opens itself, it passes victory and defeat, blessing and curse, mastery and slavery over to a historical humanity for a verdict. The dawning world brings out what is as yet undecided and measureless, and thus discloses the hidden necessity of measure and decisiveness. (p. 50)
And

By the opening up of a world, all things gain their lingering and hastening, their remoteness and nearness, their scope and limits. (p. 31)

From these statements we have developed six categories of questions to ponder the fundamental characteristics of a world, which ask about the work in an indirect way. They do not ask only about the organization itself, but about the world it does or even could or should set up. The questions are:

- What is the style of the organization?
- What is regarded as success or failure in the organization?
- What is power and authority in an organization? What grants or takes away authority and power from a group or individual?
- What does it mean for a human being to be in the organization? What is the identity, roles and reputation of a human being in an organization?
- Who do we hold ourselves as individuals and as a group accountable to?
- What is the sense of time in the organization? (How far do we look into the past or future? Are we past or future oriented?)

In order to develop the narrative we asked these questions to both senior individuals and operational level workers in a series of interviews in the aforementioned organization conducted during August and September 2005. Six members of the executive management team were interviewed by teams of two of the three researchers on the project. These interviews lasted between one and two hours. Three group interviews were held with groups of up to 20 operational level workers. These group interviews were conducted by one of the researchers, with a second researcher, and (on two occasions) the artist, in attendance. The group interviews were of two hours duration each. Thus in total we interviewed 66 members of the organization, in addition to the original briefing by the communications executive, and the discussions with the CEO. We also gathered internal organization communication and public documents, and were given access to the results of an employee climate survey and the organization strategy.

These interviews aimed to achieve an essential, phenomenological view of the organization, and therefore were intended to probe or question the very essence of the organization. As [54]Van Manen (1997) notes ‘the essence of the question … is the opening up, and keeping open of possibilities. But we can only do this if we keep ourselves open in such a way that in this abiding concern of our questioning we find ourselves deeply interested (inter-esse, to be or stand in the midst of something) in that which makes the question possible in the first instance’ (p. 43). The questions asked thus sought the
essence, not of any objective ‘truth’, but of what it was like to live in the world of the organization. In each case, ‘the question is not a problem to be solved, but a form of involvement and attunement to lived experience which animates research and writing from within’ (Friesen, Hamilton, & Cressman, 2006, p. 48).

The questions that we asked in these interviews allowed for revealing responses very often through ‘lived experience descriptions’ (Van Manen, 1997). Across both the individual and group interviews, the interviewees frequently shared their situation in an anecdotal (i.e. narrative) manner of which we depicted a number in the form of cartoons, one of which is shown in Fig. 3. A cartoon such as this is obviously an interpretation of a narrative. In this case the interviewee actually said ‘You know, once I thought I’d show them. I brought down the whole of the [x] region, just to show that I could bring it back up again. And when they asked why, I said it had just been an error, and I fixed it. But I showed them.’ The cartoon plays on this story, simplifying the notion of ‘bringing down’ the region, protecting the identity of the worker by changing the region, and introducing an element of humour with ‘Mini, would you get me another cup of coffee’. What is important in this interpretation, however, is that it has been acknowledged as essentially truthful in its portrayal of one of the major concerns of organization employees – that of not being heard or recognized.

![Fig. 3. Narrative-Visual Depicting the World of the Telecommunications Company.](image-url)
Anecdotes or narratives such as this provided minute and concrete detail of life in the organization. What was important however was not merely this concreteness (although it does give a sense of how ‘real’ things are), but rather of how this detail shows up the underlying truth of the organization. ‘In drawing us towards the concrete lifeworld experience of which they might be an example, such descriptions … help us to focus on and orient ourselves towards the question of what the experience is like, what significance it bears’ (Friesen et al., 2006). Thus, as the interviews progressed, we gained a stronger and stronger sense of the ‘truths’ of the organization – the characteristics of the world in which employees and managers found themselves. This world was characterized by grievance and a belief that employees were not being recognized or heard for their efforts. The organization itself was considered to be under constant criticism from customers and the threat of imminent competition. In the detail of the employees’ stories and in the flavour of the interviews we heard of deafness, blindness, injustice and authoritarianism. The world of the telecommunications organization was ironically characterized mostly by a lack of communication. Our task then was to develop a narrative that would both reflect these truths, and provide an alternative world, which would draw the organization towards the intended strategic reality.

The emergent process of developing this strategic narrative thus took place in the form of an interpretation rather than an explanation. The action research became, as for Grubbs (2001), a process of creative writing. The team of researchers, artist and professional storyteller met over a period of weeks to discuss (narrate) the emergent interpretation (story) of the organization as the interviews were being conducted. This process showed a direction or ‘drift’ rather than describing a final goal (Gadamer, 1961). But the drift was very clear as the process continued.

The challenge in actually creating the story was to give the people in the organization a voice rather than telling the story of the observer. A key to writing a high-quality strategic narrative is therefore to suspend one’s own interpretations and let the interviewees speak through the actual story rather than the story writer him or herself. What we sought to do was to turn the output of the interviews into a compelling story, one that would align individuals in the organization (who gave us this data) with the intended direction of the organization, as communicated by the senior executive who briefed us on the project, and in the organization strategy documentation. A story succeeds when it ‘stands out from other organizational stories, is persuasive and invokes retelling. What the story revolves around, how it is put together, and the way it is told all determine whether it becomes one worth
listening to, remembering, and acting upon. Thus strategic effectiveness from a narrative perspective is intimately tied to acceptance, approval, and adoption’ (Barry & Elmes, 1997). The appropriate, and also strategic, story needs to give a truthful account of the current world of the organization (credibility and believability) and offer the appropriate transformation that allows for the mastery of the existing challenges (novelty and defamiliarization).

In the case of the telecommunications organization what emerged in the interpretive and creative process was a number of allegories, each of which reflected the key themes of the interviews, but also provided for the possibility of mastery of the challenges that these themes presented. Allegory is ‘an extended metaphor in which authors employ narrative devices to convey a more symbolic meaning than in otherwise apparent in the text’ (Grubbs, 2001). Its use in organizational research is by no means new, as it has been adopted in several instances, probably the most well known of which is Weick’s (1996) use of the Mann Gulch and South Canyon wildfire tragedies. Our use is similar specifically, to that of Grubbs (2001) in that we not only use allegory as an interpretive lens in our research, but also in the action of communicating that interpretation for further discussion and action in the organization. Where we would suggest that our use of allegory provides a methodological development in organizational research, and in strategic management research in particular, however, is that our allegories are developed on the basis of an ontological understanding of the organization (which the allegory seeks to reflect), as a work that worlds. Therefore the allegory is useful because it is a good (truth-full) representation of that world.

In developing an allegory (which is clearly not literally true), however, it is possible that we have arguably diverted from our stated intention in adopting an ontological approach, which was to suggest a unified understanding of what an organization actually, literally, is. We have ourselves adopted a metaphor, in as much as an allegory is ‘an extended metaphor in which characters or objects in the narrative represent unique cultural attributes from their respective organizational communities’ (Grubbs, 2001, p. 380). We would suggest however, that the danger of metaphors when they are used to describe organizations is that they are so often taken literally. Metaphors, used frequently enough, can acquire the status of fact, although in an implied and assumed rather than explicit way, becoming dormant metaphors ‘through which we restrict ourselves to seeing the world in particular ways’ (Tsoukas, 1996, p. 569). For example, scientific thinking tends to suggest the notion that the corporation is a physical object, since these are the kinds of entities for which the methodologies of scientific research are most applicable. The metaphor of the machine
therefore leads managers to assume that the corporation is in fact a physical object, or at least that it will behave like one.

In contrast, in an allegory, which provides a very visible metaphor (as opposed to a ‘dormant’ metaphor such as ‘the corporation’), the essential nature of the organization as a work is not covered up, but rather revealed. The function of the allegory is to provide an interpretive lens that allows the members of the organization to see what is normally unseen, and say what is usually unsaid. The allegory does not say, the organization is this (categorically), but rather says, the world of the organization is like this (metaphorically), however without denying its very nature as being a work.

The empirical and grounded nature of the research meant that the researchers were able to gain a good – that is, credible – understanding of the world of the organization. Of several allegories that the team developed, the organization executive chose to use an allegory that reflected the reality of the organization through the metaphors of loss of speech and hearing. It tells how the elders and warriors of a tribe became, over time, unable to communicate. The strategic objective of the organization is portrayed as the re-attainment of the capabilities of speech and hearing. This allegory is ‘The Sawubona, Sanibonani Story’ which tells the tale of the Zandendaba tribe. The full allegory is reproduced in appendix.

The reader will, we hope, find the allegory interesting, but what is important about it is not that it is interesting to removed readers, or even to us as researchers, but that it is received as essentially truthful by the organization, and thus we do not offer here any sort of detailed interpretation of the story. What is more important is to convey how the story was told and received in the organization.

At the time that we developed the story, a new Chief Executive had been appointed to the organization. This story formed part of a series of roadshows at which this new CEO was introduced to the organization. At these roadshows The Sawubona Sanibonani Story was told by a professional storyteller together with a number of poems, which were read by the poet. The immediate audience response to these tellings was incredibly positive and vocal, and it is probably an understatement to say that the story was well received across the organization.

This could, of course, simply be a result of the novelty of the approach: a story of this kind is so much more engaging than another Powerpoint presentation. What we regard as more indicative of the success of the story is that the language of the allegory is subsequently being used in and across the organization to explore operational and strategic issues. Management certainly believes that the story has achieved high levels of buy-in and alignment.
Since we believe it fulfils the function of giving a truthful analysis and account of the current reality and a vivid and effective account of the future possibilities in such a way that high level of alignment among internal and external stakeholders for the purpose of realizing strategic objectives are achieved, we call the product of the action research process a ‘strategic narrative’. And since there is an overarching logic for the strategic narratives, this allows then for developing the strategic narratives in the appropriate form for different stakeholder groups (marketing campaigns, investor communications, internal communications, industrial theatre, etc.).

SUMMARY AND CONCLUSION

In this chapter we have articulated the need for increased rigour and relevance in strategic management research. We have suggested that in order to achieve both, it is necessary to consider the ontological status of the entity that strategic management is concerned with, that is, the organization. Following Heidegger’s thinking, we have developed a phenomenological ascertainment of the organization as being a work, where a work is the kind of entity that is fundamentally characterized by setting up a world for people. We have then argued that an appropriate way of giving an account of an organization is in the form of strategic narratives: narratives that give a truthful account of the world of an organization as it is and as it could be.

Since narratives play a fundamental role in human existence and are a powerful means to shaping peoples thinking and actions, we have argued that ontologically founded strategic narrative research and development meets the double challenge of strategic management research. That is, by understanding the nature of the organization, and the narratives at play in an organization we can give a rigorous account of the strategic management action in the organization, and by providing a means of developing strategic narratives themselves, we can provide relevance to strategic managers.

NOTES

1. We are not suggesting that phenomenology and pragmatism are incommensurable. Phenomenology, like pragmatism, is concerned with ‘our ways of perceiving and getting along in the world’ (Powell, 2003, p. 291). A fundamental difference however between pragmatism as espoused by Powell and phenomenology is the approach to ontology.
2. Social constructionism, for example, discusses how individuals socially construct their realities, while structuration theory asks what it is that holds social structures together over time. Discourse theory analyses organizational structures, processes, and practices. All these are very insightful ways of thinking about various aspects of organizations, as multi-dimensional constructs. Metaphorical analysis on the other hand, provides a particular and partial view of the organization, and therefore by definition speaks about entities in terms of what they are not rather than what they literally are.

3. In the tradition of ontological questioning Heidegger stands out. Heidegger has been widely acknowledged as one of the most influential thinkers of the 20th century and generations of philosophers like Arendt, Bourdieu, Derrida, Foucault, Gadamer, Habermas, Merleau-Ponty, Rorty and Sartre have acknowledged a debt to Heidegger (Dreyfus & Hall, 1992; Guignon, 1993). He, possibly alongside Wittgenstein, was one of the most influential thinkers in the 20th century philosophical discourse (Figal, 2000; Guignon, 1993). Heidegger is the thinker who is regarded as the most prominent philosopher in both asking the ontological question regarding ‘Being’ itself and the very nature of entities (Brugger, 1976). For this reason we chose him as a guiding thinker to respond to the ontological question, what is the very nature of the company? Heidegger’s work is phenomenological, in the sense of dealing with phenomena as they show up in the world. The account that follows is therefore phenomenological in its grounding and approach.

4. Humans will never be able to actually experience or fully articulate this understanding that is critical to encountering anything in the way in which it is encountered. Since world cannot be explained, yet is fundamental to being human, it remains the essential mystery of human existence that distinguishes humans from all other entities.

5. Obviously this is a hugely simplified example – we use hammers for a multitude of purposes, but this is exactly the point, that it is all the practices and entities in the referential whole, that refer to each other, and give each other meaning, within a particular world. The world of the San people, for example, as a nomadic hunter-gather world, might be expected to have no concept of a hammer as a construction tool, as we know it.

6. A community of practice is a group of individuals who engage in common practice, and in so doing, develop ‘unique interpretive repertoires’ (Boland & Tenkasi, 1995). Such a community may not coincide with organizational boundaries (Boland & Tenkasi, 1995), or even physical settings (Tyre & Hippel, 1997). A single individual may very well simultaneously be a member of multiple communities of practice, within and external to the company. Where a community of practice has a particular culture, we would link this to the notion of a world.

7. The team working on the project consisted of three researchers (Dominik Heil, Louise Whittaker, & Shupi Heil), an artist (Alistair Findlay ) and a professional organizational storyteller (Peter Christie).

8. ‘Sawubona, Sanibonani’ is the Zulu greeting in the first person singular, and first person plural. It literally means, ‘I/we see you’.

9. ‘Zanendaba’ is a pun. An ‘Indaba’ is a meeting. ‘Zandendaba’ implies ‘those who talk’.

10. An African philosophy of existence which holds that ‘a person is a person because of other people’.

REFERENCES


APPENDIX. THE SAWUBONA, SANIBONANI STORY

If I See You, And You See Me,
Together We’ll The Richer Be.

(An Intimacy, ‘Into Me, See’ Story.)

Just as night fell, the Zanendaba hunters crawled exhausted into a small cave half-way up a hill on the edge of their fields, still some distance from their village, seeking safety from a ferocious band of marauding rival tribes. The Zanendaba hunters had earlier been out hunting, oblivious to any danger, for if the truth be known, no-one had, since almost their very beginnings, dared to attack their great and powerful tribe, the surrounding bushveld over which they enjoyed almost complete hegemony.

But that afternoon they at first heard the sound of warlike drumbeats, and then saw the distant flames of enemy fire torches. Although the Zanendaba hunters beat a steady retreat towards their village, unprepared for battle, the enemy tribes were hungrier and more athletic, making greater ground than their prey, and the Zanendaba hunters were forced to seek refuge in the cave. On entering the darkness they lay down breathlessly, then scrambled around on all fours looking for some rocks with which they could protect the cave’s entrance. At no time did a word pass between their lips, not
entirely from fear of detection by their enemies, but for their belonging to a tribe entirely and utterly cursed by its ancestors.

Where once and always the Zanendaba tribe had roamed free across the entire bushveld, its members had grown fat and arrogant, neglecting sometimes even to perform the rituals and tell the stories from which the blessings of their ancestors arose. They squandered their abundance, neglected to do their work, and preferred feasting and festivities. But most of all, they stopped caring for each other – stopped seeing one another in the right light – and nothing angers the ancestors more than a tribe that has lost its spirit of ubuntu. The elders would have a lot to say to the hunters, with instructions and orders and expectations and dictates, but seldom listened in return, and the hunters would often heap scorn and ridicule the elders, maliciously gossiping behind their backs. One night, when all the tribe members were gathered for yet another celebration, a tremendous storm blew up, and amidst crashes of thunder and cracks of lightning, a demon burst out from the flames of fire, cackling into the suddenly still night.

‘Oh, elders, your mouths are so full of air, I declare that your ears be filled with despair’. As the demon cried out his incantation, so all sound and noise disappeared from the ears of the elders, all to remain completely deaf from that day since. ‘Oh, hunters, your tongues so full of bile, so too your mouths I shall beguile’, the bad and bold apparition barked demonically. In an instant the Zanendaba hunters and their wives, every man and every woman, were struck completely mute, to remain voiceless forever more. Then the demon spat upon the flames, dousing completely the fire, the embers within which he disappeared as suddenly as he had appeared.

So it was that many, many moons back the Zanendaba tribe had been cursed to have elders who could not hear, and hunters who could not speak. At first this strange affliction did not seem to overly bother the Zanendaba, who carried on with their lives as best they could. However, word had slowly spread to rival tribes in their vicinity, who sensed their old enemy’s new vulnerability. It took a very long time but eventually the warlike tribes had mustered the strength and agreed the strategy upon which events might turn in their favour. Indeed, nothing would wake the Zanendaba from their slumber quite like what befell them on that fateful day when, in a strategic alliance, the marauding tribes attacked, and the Zanendaba hunters, as we know, found themselves ensconced in the cave high up on the hillside.
Huddled together, in fear of their very lives, the Zanendaba hunters looked across the valley towards their village, from where deep inside they could remotely hear the noisy, frenzied and unintelligible cackling, commanding, crying, cursing and cavorting of their elders, naturally none of whom could hear the commotion they were making, and blissfully unaware of their enemy’s intent.

All the while however, from the cave the Zanendaba hunters could hear their enemies’ war drums and war chants, and see the flames of their burning stakes as the enemy warriors approached the village. Although distant at first, the noises quickly became louder, and the flames brighter. In desperation, the Zanendaba hunters banded together to try and chant a warning to their elders, but not a sound was heard to come out of their mouths. In a panic, they jumped around like crazy banshees, beating on their chests and stamping on their feet, raising a cloud of dust that threatened to suffocate all including the ancestors, slapped and licked wildly at one another, but all this was to no avail. Without voices, they could not draw the attention of their elders to their impending plight.

Meanwhile, in the village itself, the Chief of the Zanendaba and a few elders went to the edge of the village in expectation of providing the returning hunters with a warm welcome from their hunt. Although still low in the sky, the moon was bright and standing alone on the edge of the bushveld, the Chief and his elders could see across the horizon, above the silhouettes of the thorn bushes and the fever trees. Then they saw the flames of the enemy fire torches, and mistaking them for those of the returning hunters, went forth into the clearing, unable to hear the battle drumbeats, and the warning rhythms of conflict and conquest.

From beneath the cover of darkness the leading enemy tribe suddenly attacked, immediately ensnaring the Chief of the Zanendaba and his elders. Hordes of enemy warriors poured forth into the kraal, loosing the cattle, setting alight some of the village’s outer huts, capturing some of the Zanendaba women, before hastily retreating into the bush, and disappearing into the African night. On witnessing the mayhem happening in the village below, the Zanendaba hunters hurriedly clambered down from the cave, and sprinted to protect the remaining homesteads. But this tactic was already far too little and far too late. By the time they reached the village, the marauding enemy tribes had already long fled.

The Zanendaba hunters made their way to the Royal Kraal, where the surviving elders were fearfully and fretfully trying to assess the extent of the death and destruction. However, not having witnessed what had happened
with their own eyes, with all the conflict having occurred on the village’s periphery, the surviving elders mistakenly believed that the village had been betrayed and plundered by its own hunters, seeking to establish their own Kingdom. What else, they thought, could explain the disappearance of the Chief, some elders and some of the women, and their own hunters appearing unblemished and unhurt? Try as the hunters might to explain the circumstances of the marauding enemy tribes to the elders, the former could not speak and the latter could not hear. It was impossible to achieve understanding between the two. So the elders ordered the death of those they saw as the hunter’s ringleaders, the punishment being tragically enacted before first light. A gloom descended upon the Kingdom of the Zanendaba, which lasted for many seasons, during which time not a sound was uttered, not a word was spoken. Elder and hunter alike churned over in their own dismay. Together but apart, the two camps brooded from within.

Then one day at the burst of spring, a wise elder and a wise hunter sat together for the first time in many months. Nothing needed to be said, nor anything to be heard, from one to the other, for at a level far deeper than language, they began to really communicate. Intuitively and instinctively they began properly to see, a sense they had never lost. In agreement with the other elders and other hunters, they slaughtered the village’s oldest and heaviest ox, and sacrificed it to the ancestors that night. For an entire month the tribe fasted, with nothing but air passing their lips.

Then the wise elder and the wise hunter convened a ritual hunt, and for the first time in the Zanendaba’s long history, since pre-antiquity, both elder and hunter hunted together. A single arrow in the left eye felled Arrogante, the giant elephant, greatest of the bushveld’s vast elephant herds, and whilst some of tribe prepared the carcass for sacrifice, the ritual hunt continued. As it did so, something marvellous, miraculous and magical happened, a mystical alchemy unfolded.

The elders found that, in the company of the hunters, the less they spoke the more they seemed to hear. The hunters, on the other hand, discovered that the less they heard, the more they themselves were able to speak. And by the time the long ritual hunt had finished, and the Zanendaba had returned to their village, elders and hunters alike could both speak and hear, and what was said was heard.

Sitting around the fire later that evening, over the cauldron, and having sacrificed Arrogante to the ancestors, the wise elder and the wise hunter looked deeply into each other, knowing that only when the eyes see can the mouth speak and the ears hear. Then they began to drum and they began to chant, quietly and alone at first, then louder and louder, and together with
all their brothers and sisters, elder and hunter alike. The spirit of ubuntu had been resurrected in the Zanendaba, and the tribe sang out their praises.

If I see you, and you see me, together we'll the richer be.
If I see you, and you see me, together we'll the richer be.
If I see you, and you see me, together we'll the richer be.

So thereafter for the Zanendaba tribe, with the spirits of the ancestors appeased, never were the yields more plentiful, the hunts more bountiful and the times more prosperous.

By Peter Christie
This page intentionally left blank