Journal of Management Studies 43:8 December 2006
0022-2380

POINT COUNTERPOINT

Encouraging Best Practice in Quantitative Management Research: An Incomplete List of Opportunities*

Raj Echambadi, Benjamin Campbell and Rajshree Agarwal

College of Business Administration, University of Central Florida;
The Wharton School, University of Pennsylvania;
College of Business, University of Illinois at Urbana Champaign

ABSTRACT The paper identifies some common problems encountered in quantitative methodology and provides information on current best practice to resolve these problems. We first discuss issues pertaining to variable measurement and concerns regarding the underlying relationships among variables. We then highlight several advances in estimation methodology that may circumvent issues encountered in common practice. Finally, we discuss approaches that move beyond existing research designs, including the development and use of datasets that embody linkages across levels of analysis, or combine qualitative and quantitative methods.

INTRODUCTION

Social research comprises two broad methods of logical reasoning: (a) deductive reasoning that involves the confirmation of hypotheses from theories; and (b) inductive reasoning that involves the development of generalizations from specific observations (Kerlinger, 1973). For example, Christensen (2006) highlights the importance of both inductive and deductive reasoning in the development of a unique theory of disruptive innovations. Since the process of management theory building benefits from the application of both approaches, we are pleased that this article on quantitative (deductive) methods is accompanied by the article by Shah and Corley (2006) in the area of qualitative (inductive) methods.

Based on our experiences as authors, reviewers, and readers of academic journals in the strategic management area, we recognize a need for extending the quantitative methods in the management research toolkit. The advancement of the field rests on well

Address for reprints: Raj Echambadi, Dept of Marketing, College of Business Administration, University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816-1400, USA (rechambadi@bus.ucf.edu).

tested theory, which requires addressing frequently encountered methodological errors, implementing better techniques and broadening of the scope of quantitative research. To this end, we identify some limitations and errors that recur in quantitative management work and provide information on current best practice to resolve these limitations. Readers should consider this paper a primer that identifies and presents solutions to common research problems. Our review of each solution is necessarily brief; interested readers should read the cited sources for enhanced clarity.

The paper is organized in the format similar to the empirical section of academic papers. We first discuss issues pertaining to variable measurement, and then address concerns regarding the underlying relationships among variables (i.e. causality and endogeneity). We then turn to estimation and highlight several advancements in estimation methodology that may circumvent issues encountered in common practice. Next, we discuss issues concerning interpretation of results. In the final section, we discuss approaches that move beyond existing research designs, including the development and use of datasets that embody linkages across levels of analysis or combine qualitative and quantitative methods.

VARIABLE MEASUREMENT: ILL-DEFINED CONSTRUCTS LEAD TO POORLY TESTED THEORY

Constructs comprise the basic building blocks that connect theory development to testing. Boyd et al. (2005a) caution that poor construct measurement is a serious threat to management research. Errors in measurement of a single independent variable can dramatically bias any or all of the other coefficients, thus leading to both Type I and Type II errors in hypothesis testing (see MacKenzie, 2001). For example, Boyd et al. (2005a) show that measurement error is a primary cause of the divergent findings in diversification research.

Measurement error is addressed differently in the psychometric and econometric literatures, in part due to the differences in research design and data collection methods. Psychometric approaches to measurement error can benefit management research primarily through ex-ante accommodation during the primary data collection stage, while econometric techniques allow for ex-post correction for measurement error in secondary data.

Ex-Ante Accommodation in Construct Measurement

When collecting primary data (e.g. surveys), researchers need to address issues of construct validity and reliability. Well-defined construct definition is a necessary but not a sufficient condition for construct validity. After appropriately defining constructs, researchers must focus on developing measures that adequately capture the entire domain of the construct. While many management research studies utilize multiple items to measure their constructs, measures of complex constructs using single items are unfortunately still quite prevalent (see Boyd et al. (2005b) for problems with single item measures). We urge researchers to use multiple items to obtain correct estimates of relationships and also to minimize Type II errors. If multiple items are unavailable, it is imperative to generate a set of items tapping the domain of the construct and pretest the items rigorously before use in the field. It is also extremely important for researchers to think about 'measurement relations' or how the multiple items relate to their constructs (MacKenzie, 2003).

Additionally, when choosing variable measures, it is important to delineate between reflective and formative measures, and use appropriate statistical tools to model these constructs to avoid construct misspecification. Most management studies use measures that are reflective in nature; i.e. the unobserved, latent variable is reflected in the measures being used. For example, job satisfaction can be measured using global items such as 'I am satisfied with my job', and 'In general, I like working here' (cf. Mitchell et al., 2001). These two items are interchangeable (cf. Nunnally and Bernstein, 1994), and co-vary due to the underlying job satisfaction construct. Yet, a common (and often serious) mistake encountered in management research is the use of formative indicators in a reflective setting. Formative indicators, first introduced by Blalock (1964), are measures that cause the change in a latent variable. In our job satisfaction setting, formative items may include individual satisfaction questions pertaining to pay, promotion, supervision, contingent rewards, co-workers etc (Spector, 1997). In this formative case, the items are not interchangeable and the scores on the individual items drive the overall job satisfaction. Omission of any one formative indicator may alter the nature of the construct itself (Diamantopoulos and Winklhofer, 2001). Treating formative measures as reflective measures may lead to invalid estimates (Chin, 1998).

Yet another error relates to the use of reliability measures to test dimensionality of constructs. For reflective measures, it is important to report reliability measures such as Cronbach's alpha and composite reliability (Fornell and Larcker, 1981). Large values of reliability, say values of 0.7 or more, indicate that the sample of items perform adequately in capturing the domain of the construct. If the reliability values are low for reflective measures, then the items with the lowest item-to-total correlations may be dropped (Churchill, 1979). Too often, however, we see that researchers report high alpha values as proof of unidimensionality of their constructs. This practice is incorrect. Cronbach's alpha is a measure of reliability, not of dimensionality. To test dimensionality, factor analysis approaches must be employed.

For formative measures, traditional measures of reliability are not appropriate (Diamantopoulos and Winklhofer, 2001; Jarvis et al., 2003), yet researchers often drop formative items with the lowest factor loadings to increase the reliability of the constructs. Such faulty practice changes the empirical meaning of the latent construct and undermines construct validity (MacKenzie, 2003). We urge researchers to follow the advice of Boyd et al. (2005a) that authors place great emphasis on reliability and validity checks to ensure construct validity.

Ex-Post Correction of Measurement Error

As pointed out earlier, multiple items per construct may be used to partial out measurement error using structural equation models. Alternatively, one can 'fix' reliability estimates of a single item measure based on established scales with known reliabilities in a structural equation modelling framework (Bollen, 1989).

From an econometric perspective, measurement error in an independent variable may cause it to be correlated to the error term in the estimation equation (Greene, 1997). In such cases, instrumental variables can correct for the bias introduced by measurement error in the regression estimates (Wooldridge, 2002).[1] An instrumental variable is a variable that is correlated to an independent variable (*instrument relevance*) but uncorrelated with the error terms (*instrument exogeneity*) and thus can be used to correct for measurement error and to enhance construct validity.[2]

Instrumental variable estimation is implemented as a two-stage least-squares approach (2SLS). In the first stage, the instrumental variables are used to predict the variable(s) that are measured with error. In the second stage, the estimated values from the first stage are used in place of the independent variables (Greene, 1997).[3] Since the predicted variable values are no longer correlated with the error term of the dependent variable, the measurement error issue is addressed. Care must be taken to ensure that there are enough exogenous variables in the data so that each instrumental variable can be uniquely identified (Zohoori and Savitz, 1997).

It is also important to remember that all instruments are not created equally. Weak instruments, i.e. instruments that do not predict much variation in the relevant dependent variable may lead to unreliable inferences in the second stage. The F-statistic for the first stage of an instrumental variables model should be included in the tables; in typical cases, a value of less than 8.96 is considered to be weak,[4] and such instrumental variable estimates will lead to Type I and Type II errors in hypothesis testing (Stock and Yogo, 2002). In the presence of weak instruments, one could implement estimation techniques that are more robust to weak instruments (see Stock et al., 2002 for an overview of alternative estimation techniques).

DEMONSTRATING THE RELATIONSHIPS BETWEEN VARIABLES

Well-designed tests of hypothesized relationships need to address and potentially reject issues related to alternative relationships among the relevant variables. We now turn to concerns related to causality and endogeneity that may limit the scope of empirical claims.

Causality

While causality can be demonstrated using panel data and experimental data, a common criticism of empirical work using cross-sectional data is that the results do not demonstrate causality. Researchers must first theoretically motivate the causal relationships between variables and then look for empirical evidence of the causal relationships. The presence of non-zero effects in the absence of a strong theory does not demonstrate causality. Causation can only be demonstrated when the following three conditions occur: (1) concomitant variation between the variables of interest; (2) evidence of clear temporal ordering of the variables; and (3) when all other spurious influences are controlled (Cook and Campbell, 1979). The most common approach is to address these issues through an exhaustive approach. Researchers can provide evidence of one-way causality in support of the model by providing robustness checks that: (1) reject reverse causality; (2) demonstrate the elimination of omitted variable bias; and (3) ensure that the correlations are robust to different specifications and samples. This approach relies on convincing the reader that theorized one-way causality is the only remaining option.

Endogeneity

Endogeneity occurs when the independent variable included in the model is correlated with the error terms. Ignoring endogeneity may lead to biased and inconsistent estimates. Sources of endogeneity include reverse causality, simultaneous causality, and omitted variables.

To begin with, researchers can check whether endogeneity is a valid concern by using the test developed by Hausman (1978). The Hausman test for endogeneity compares whether least squares and instrument variable estimates of the model are statistically different from each other. If there is no endogeneity in the model, both coefficient estimates are consistent and their difference converges to zero, while endogeneity results in the inconsistent least squares estimates being significantly different from the consistent instrumental variable estimates.

Reverse causality can be ruled out easily if researchers have access to panel data or can design experiments to isolate causality. With panel data, if there is a significant relationship between the dependent variable and a lagged independent variable, but no significant relationship in the other direction, evidence for the causal relationship is clear (Granger, 1969, 2003). If endogeneity is suspected because of an omitted variable, the problem is easily rectified if an adequate measure for the omitted variable can be included.[5] However, in most cases, we do not have data on the omitted variables. Ignoring the problem, though, results in an omitted variable bias. The use of panel data and estimation of fixed effects/first differencing in the presence of time-constant omitted variables may provide a solution in such a situation (Wooldridge, 2002). A common econometric solution to endogeneity arising out of omitted variables in cross-sectional data or in the presence of time varying omitted variables is to utilize an instrumental variable approach.[6] Associated with this omitted variable bias is the notion of unobserved heterogeneity: unmeasured, non-random differences across observations that are usually captured as part of the error term in a regression model (Berg and Mansley, 2004). Zohoori and Savitz (1997) suggest that instrumental variables may be used to eliminate the confounding effects created by unobserved heterogeneity in the presence of endogeneity.

Another strategy to effectively eliminate endogeneity concerns is the use of an experiment. Classical experiments where subjects are randomly assigned to experimental conditions allow researchers to control the effects of extraneous variables and, in doing so, isolate the true effects of manipulated (or treatment variables) variables on the dependent variables (cf. Boland et al., 2001). When laboratory experiments are not feasible or appropriate,[7] researchers may be able to rely on quasi-experiments, a research design having most of the characteristics of a classical experiment without the random assignment of subjects to the experimental conditions (Cook and Campbell, 1979). For example, in investigating the impact of entrepreneurial firm status on an individual's earnings potential, Campbell (2006) compares the earnings of individuals joining an entrepreneurial firm (treatment group) to a matched sample of individuals joining an established firm (control group), while accounting for similarities in other dimensions. In addition to this straight matching approach to develop a quasi-experiment, other methods include propensity-score matching (Dehejia and Wahba, 2002) and regression discontinuity analysis (Ashenfelter and Card, 1985; Chay et al., 2005).

Any discussion of endogeneity will be incomplete without mentioning a common mistake made by authors and reviewers in terms of conflating sample selection bias and endogeneity bias. The two biases stem from different sources; while endogeneity questions the exogeneity of an explanatory variable, sample selection bias occurs when the dependent variable is observed only for a restricted, non-random sample. The latter can be rectified using Heckman's selection correction model (1979). This point should not be construed to imply that endogeneity and sample selection biases may not occur together in the same model. If they do occur simultaneously, a more sophisticated modelling approach, as discussed in Amemiya (1985), may be used to account for both these biases.

CAPTURING INTERACTIONS

As management theory moves beyond hypothesizing main effects of the explanatory variables and towards unearthing contingency conditions, the use of moderated models is becoming more pervasive. In essence, the interaction (or moderator) effect suggests that the effect of an independent variable, x_1, on the dependent variable, y, varies depending upon a third variable, x_2. The interaction in a moderated model is estimated by including a cross-product term as an additional independent variable:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_0 + \beta_c x_c + \varepsilon, \tag{1}$$

where x_c plays the role of other covariates that are not part of the moderated element. The interaction term, x_1x_2, is likely to be correlated with the term x_1 and this correlation has been interpreted as collinearity.

Mean and Residual Centring in Interaction Models

The common practice of mean-centring of variables x_1 and x_2 prior to creating the interaction term (see Aiken and West, 1991) to reduce correlations between the independent variables and the interaction terms has been recently called into question. Although the magnitude of the correlations does indeed decrease for normal random variables,[8] Echambadi and Hess (2007) analytically show that the mean-centred models are mathematically equivalent to models that use raw (or uncentred) terms. As a result, collinearity is not alleviated by mean-centring. In a similar vein, Echambadi et al. (2006a) show that the residual-centring approach proposed by Lance (1988) to mitigate collinearity problems leads to uninterpretable simple effects. Therefore, we recommend *against* the use of mean- or residual-centring variables for collinearity reasons.

A casual content analysis of the premier management journals reveals that scholars commonly use bivariate correlations to assess the presence of collinearity in moderated models. However, bivariate correlations are neither necessary nor sufficient conditions of multicollinearity (Ofir and Khuri, 1986) and reliance on a single diagnostic may provide misleading results (Mela and Kopalle, 2002). To address this issue, Echambadi and Hess (2007) recommend that researchers use and report multiple diagnostic tools to assess potential collinearity problems. Reliance on multiple collinearity diagnostics may reduce false alarms about the presence of collinearity. The effects of multicollinearity are indistinguishable from the effects of micronumerosity, or small sample sizes (Goldberger, 1991), and hence increasing the sample size can mitigate the loss of power associated with collinearity (Mason and Perreault, 1991).

Main, Simple and Interaction Effect Models

We highlight three prevalent misconceptions in the context of modelling interaction effects. First, Irwin and McClelland (2001) point to the erroneously held belief that the change in the magnitude of a simple effect when an interaction term is added is due to the collinearity introduced between the main effects and the interaction term. As a result, researchers first estimate and interpret a 'main effects only' model and then estimate a 'full' model that includes the interactions. However, the change in simple effects that occurs when the interaction term is added is because the 'main effects only' model and the 'full' model are inherently different models. In a 'main effects only' model, the coefficient, say β_1, represents the average effect of x_1 across all levels of x_2. However, in the interaction or 'full' model that employs raw terms, β_1 represents the effect of x_1 when x_2 equals zero. The differences in the coefficients across the two models thus have little to do with the collinearity introduced by the interaction term in the 'full' model. If the 'full' model represents the theoretically justified model, then estimating and interpreting a 'main effects only' model will constitute a theoretical misspecification and the coefficients of a 'main effects' model will suffer from omission bias (Echambadi et al., 2006a). Therefore, we recommend that researchers estimate both the simple effects and the interaction effects simultaneously in a single model.

Second, management researchers do not always include all the simple effect terms in a moderated model. For example, in equation (1), if x_2 is not added to the full model, the significance of x_1x_2 is confounded with the omitted simple effect of x_2 and as such the interaction term becomes uninterpretable (Irwin and McClelland, 2001). Third, many researchers split continuous independent variables into categorical variables in order to reduce collinearity concerns. Irwin and McClelland (2003) show that splitting continuous variables into categorical variables leads to reduced power and deleterious consequences. If the variable must indeed be dichotomized, then researchers must conduct robustness tests to show that results do not change when the dividing line is specified differently; for example, use of a median split instead of a mean split.

BETTER-PRACTICE ESTIMATION METHODOLOGIES FOR STRUCTURAL MODELLING

Structural Equation Modelling or Partial Least Squares?

Structural equation modelling (SEM) and partial least squares (PLS) techniques have gained immense popularity in the management field in the last decade, in part due to their inherent abilities in testing complex theoretical structures. SEM is a largely confirmatory, covariance-based approach wherein the focus is on minimizing the differences between the covariances derived from the empirical data from those predicted by the model-implied covariance matrix. Therefore, goodness of fit between the predicted covariance and the observed matrices is assessed to identify a suitable model. In contrast, PLS is a component-based exploratory approach wherein the latent variables are estimated as exact linear combinations of the observed measures (Chin, 1998). Since the focus of PLS is on explanation of variance, measures of R^2 are used to identify a suitable PLS model (Hulland, 1999).

When should one use PLS or SEM? Under the joint condition of large sample sizes (consistency) and large number of indicators per factor (consistency at large), PLS estimates of the factor loadings and structural coefficients approximate that of the SEM estimates (Barclay et al., 1995). The use of SEM or PLS should be dictated by the goal of the research. SEM is the obvious choice for confirmatory theory-testing research. Alternatively, if the goal is to maximize the variance of manifest variables in exploratory situations, then PLS seems more appropriate (Chin and Newsted, 1999). Also, early notions were that PLS is more appropriate to model constructs measured with formative indicators (Barclay et al., 1995), and that the SEM approach in such cases is fraught with identification problems that have been difficult to work through (Chin, 1998; MacCallum and Browne, 1993). Recently, however, Jarvis et al. (2003) show ways of achieving identification in models using formative indicators in SEM.

A cursory examination of the SEM papers in our field reveals that many scholars mistakenly assume that a 'good fitting model' is the final goal of any SEM modelling exercise (cf. Chin, 1998). Because the underlying sampling distributions of these goodness-of-fit indices are unknown, the appropriateness of any given fit index for a given setting may be problematic (see Shook et al., 2004 for their critique of SEM studies in strategic management). Also, it is possible to obtain high goodness-of-fit measures for models with poor factor loadings (Chin, 1998). Furthermore, pure reliance on model fit also ignores effect sizes (Cohen, 1990). Therefore, we recommend that researchers consider a combination of diagnostics: multiple fit indices (Shook et al., 2004), substantial factor loadings, and sizable structural coefficients, to choose the best model (Chin, 1998). Specifically, we suggest that researchers use 0.7 as a suggested cut-off for factor loading of established constructs (Hulland, 1999)[9] and we follow Chin (1998) and Meehl (1990) who suggest that standardized paths should be around 0.20 to be considered practically meaningful.

Under the belief that PLS makes modest demands on data, authors across disciplines are increasingly estimating PLS models on inappropriately small sample sizes (cf. Marcoulides and Saunders, 2006). While it is true that appropriate sample sizes tend to be smaller for PLS (a limited information estimation procedure) than for SEM (a full information procedure), Chin and Newsted (1999) point out that smaller structural path coefficients in PLS (e.g. 0.20 or below) do not obtain statistical significance until large sample sizes, say 200 cases, are achieved. Therefore, it is important that scholars not use PLS as a panacea when dealing with small sample sizes. Researchers should examine the stability of the estimates and the magnitude of the standard errors since unstable coefficients with large standard errors are usually an indication of inadequate sample size (Marcoulides and Saunders, 2006).

Testing Interactions in Latent Variable Models

A multi-group approach is most commonly used to test interactions in SEM. In this approach, the moderator variable is typically categorical. Separate SEM models are estimated for each level of the categorical variables, and then compared using the traditional measures of goodness-of-fit statistics and the statistical significance of the coefficients. Given complexities in estimating continuous variable interactions, some researchers attempt to simplify the problem by making continuous variables categorical. However, splitting continuous variables leads to a loss of information (Irwin and McClelland, 2003) and hence we recommend that researchers test moderating effects by creating and testing multiplicative terms of the component variables (see Cortina et al. (2001) for details of various approaches).

Similar to SEM models, interaction effects can also be easily tested in PLS. Chin et al. (2003) propose a way to test interaction effects using PLS (see Sarkar et al. (2001) for use of PLS to test latent variable interactions in strategic management). Analogous to multiple regression, the multiplicative 'latent interaction variable' is reflected by indicators created by multiplying all possible products from the predictor and the moderator variables. The PLS procedure is then used to estimate the latent variables and obtain the moderating effects. Chin et al. (2003) also highlight an approach to test interaction terms created from constructs measured with formative items.

Hierarchical Linear Modelling

Hierarchical data structures refer to inherent nesting of data within a macro structure. For example, employees of a firm exist within organizational units that are nested within the firm. Put differently, the data exist in multiple levels: individual-level, unit-level, and firm-level. As a result, employees within units tend to be more similar to each other and are likely to be different from employees from other units because employees within a unit are assigned deliberately due to certain factors (e.g. skill and education levels, job specialization, etc). In cases wherein multiple employees within a unit are relatively homogeneous on certain dimensions, the random error component within such nested data will also include an aggregate level random error, thereby violating the statistical assumption of independence of observations (Bryk and Raudenbush, 1992). Also, this aggregate level random error may vary across groups thereby violating the homoskedascity assumption (Bryk and Raudenbush, 1992; Hofmann and Gavin, 1998). Failure to address these violations while estimating may lead to biased estimates, smaller standard errors, and incorrect statistical inferences (Bryk and Raudenbush, 1992).[10]

Hierarchical Linear Modelling (HLM), also known as multilevel modelling, is a random coefficient model that is appropriate for modelling nested data (Hofmann, 1997). For example, consider individuals (Level I) that are part of the same firm (Level II). At Level I, individual level employee outcome is regressed onto the individual level predictors. The resultant parameter estimates from the Level I (i.e. intercepts and slopes) regressions are then used as outcome variables in the firm (Level II) analysis where they are modelled as a function of firm-specific (Level II) variables. Level I parameters are allowed to vary across groups and the variance and covariance of the Level II residuals are separately estimated (cf. Bryk and Raudenbush, 1992).

HLM offers a number of advantages to the researchers. If the goal of research is to disentangle the unique effects of different variables at multiple levels or to model the interactions of variables across multiple levels that allows for more interesting research questions, then the use of hierarchical linear models (HLM) becomes imperative. If variables exist in multiple levels, HLM enables researchers to obtain unbiased estimates of the variables at all levels and accurate standard error estimates (Bryk and Raudenbush, 1992). Also, by virtue of accounting for clustered observations, it improves the efficiency of the estimates. Researchers can scale the Level I independent variables in one of three ways: uncentred, grand mean-centred, or group mean-centred. The type of scaling employed should be chosen with great care as it influences the interpretation of the parameters (cf. Hofmann and Gavin, 1998).

BETTER-PRACTICE ESTIMATION METHODOLOGIES FOR LONGITUDINAL DATA

While longitudinal data techniques are typically more powerful than cross-sectional studies, there are challenges in modelling longitudinal data as well. A failure to incorporate the correlations that occur in the data by definition may lead to incorrect estimation of regression model parameters (Ballinger, 2004). Broadly, there are two major means of analysing longitudinal data with correlated responses: generalized linear mixed models (GLMM) that extend the general linear model by allowing estimation of both fixed and random effects, and generalized estimating equations (GEE) models (Allison, 2005).

When analysing correlated longitudinal data, the choice of the approach (GLMM or GEE) to be used depends on the purpose of the study. GLMMs are commonly used to obtain subject-specific effects, while GEE models are appropriate for estimating population-averaged coefficients (Hu et al., 1998). A subject-specific coefficient is an estimate of what would happen to a specific individual observation when one unit of the independent variable is increased, while a population-averaged coefficient is an estimate of the change in the average response for the entire population of observations for a change in one unit of the independent variable (Allison, 2005; Ballinger, 2004). In other words, research focus dictates the type of approach employed. For example, when the survival response for an individual firm is of interest, GLMMs are appropriate. On the other hand, if the interest is on the differences in survival responses between two cohorts of firms that entered at the same time period, then population averaged estimates are more appropriate.

The GEE approach utilizes quasi-likelihood estimation and models the covariates of interest without explicitly accounting for individual heterogeneity whereas GLMMs utilize full-likelihood methods to estimate the covariates after accounting for heterogeneity (Zeger et al., 1988). As a result, GEE models are not appropriate for situations wherein variance and/or covariance parameters are of significant interest (Hedeker and Gibbons, 2006).

Longitudinal SEM

SEM approaches can also be used on longitudinal data.[11] Beyond the obvious advantages of modelling causal relationships, SEM enables researchers to compare competing models of causal relationships (Farrell, 1994). Also, SEM approaches can be used to model growth over a period of time. Researchers have argued for the superiority of SEM growth modelling over other analytical techniques because it allows more flexibility in testing alternative models of growth over time (Sivo et al., 2005).

TYING RESULTS TO THEORY

Inferences Regarding Causality

A common trap in interpreting statistical results is to claim causation when the empirical methodology supports correlation. As discussed in the earlier section, from an empirical perspective, scholars should address alternative explanations such as reverse causality, causality driven by an omitted variable, and coincidental correlation. Correlational results alone cannot be taken to imply support of causal mechanisms. Quantitative researchers need to be very careful with how they design their studies, interpret their empirical work, and discuss their results. Often, phrases like 'the data are consistent with the theory', or 'the empirical analysis does not reject the theory' are much more accurate (if less dramatic) than 'the data support the theory' or 'the data prove the theory'.

Interpreting Results in Linear Models

In interpreting results, it is extremely important to differentiate between statistical significance and economic significance. Statistical significance is only a test of the data; it is dependent on the number of observations in the data. Economic significance, on the other hand, asks the question: 'is the observed effect large enough to have a meaningful impact?' Calculation of economic significance is highly context-specific. Depending upon the size of the data, researchers must bear in mind that it is possible to find a number of statistically significant coefficients that are practically meaningless with little impact on the dependent variable. Framing and discussing the results in terms their economic significance will permit readers to gauge their theoretical and managerial importance.

Interpreting Results in Non-Linear Models

Unlike linear models, the sign and significance of the estimated coefficients in non-linear specifications (e.g. limited dependent variables), do not capture the whole story because the effect of any one independent variable is conditional on the values of the other independent variables, as well as the rest of the parameters in the equation. Therefore, the correct magnitude of an independent variable is given by its marginal effect (how much does a one unit change in an explanatory variable impact the dependent variable). These marginal effects are computed by setting the other variables at some specific value (Long, 1997).[12] Since the marginal effect changes across the distribution of observations, researchers have to choose how to present the marginal effects. Common approaches include the often confused 'average marginal effect' and the 'marginal effect at the average'. The 'average marginal effect' is created by calculating the marginal effect for each observation in the data and then taking the average across all observations (Hoetker, 2007). The 'marginal effect at the average' is calculated by taking the average value of each variable across all observations and computing the marginal effect at these average values.[13] A good practice approach is to identify theoretically meaningful values of the explanatory variables, and compute the marginal effects for these values or to present graphs and charts to capture the dynamics of the marginal effect (Hoetker, 2007). Regardless of which technique is employed, better practice entails reporting not only the coefficients, but the marginal effects of the variables at values of the variable that are of theoretical interest, which may extend beyond simply the mean values.

CONCLUDING THOUGHTS

In this paper, we have identified some of the most often encountered problems in quantitative management research. Table I provides a summary of these concerns, their consequence and recommendations for solutions. While not an exhaustive list, we hope that the paper provides readers with better practice solutions to these issues. Given our focus on prevalent quantitative methodology issues, we concentrated on existing problems in commonly used methodologies. There is great value, however, in moving beyond the popular research design. In our concluding section, we highlight a few examples of some under-utilized research designs that can help address critical issues in management theory.

Time-Series Research Design

Time series techniques may be used to study and/or disentangle short-run and long-run effects of strategic actions in both stable and evolving conditions. For example, innovation research suggests that the number of firms increases rapidly during the growth stages of an industry, while the mature stages are characterized by stability and a few large firms dominating the industry (cf. Agarwal et al., 2002). The latter period is best modelled using stationary processes that assume that probability distributions are stable over time and that the means and/or covariances are equal over time. Evolving markets, on the other hand, are characterized by non-stationarity and require special time series approaches (Hamilton, 1994). Further, even in stationary scenarios, time series analyses can be fruitful in analysing outcomes of actions based on game-theoretic assumptions of competitor reaction. As Dekimpe and Hanssens (1999) show, strategic changes undertaken by firms may create either a sustained change into the future (hysteresis) or may provide at best short-run performance gains since the long-run performance metrics revert back to the underlying stationary process (business as usual). Thus, time series data can be fruitful for uncovering new insights, particularly when continually changing market conditions make it difficult to relate current action to future performance (Dekimpe and Hanssens, 1995).

Table I. Common problems, their consequences and recommended solutions

Problem/error	*Consequence*	*Solution*
Measurement error		
Single item measurement	Biased estimates; attenuation of coefficients in a simple regression; under- or over-estimation of coefficients in a multiple regression.	Use multiple items wherever possible; if not possible, use instrumental variables to correct for measurement error. Alternatively, fix the reliability estimates of single item measures.
Confusion between formative/reflective measures	Invalid estimates due to construct misspecification; inappropriate use of reliability indices in the case of formative measures.	Clearly specify the nature of relationship between the manifest items and their constructs. Do not use measures to diagnose reliability problems in the case of formative measures.
Use of weak instruments	Invalid inferences due to Type I/II errors.	Use stronger measures as instrumental variables; use estimation techniques that are more robust to weak instruments.
Relationships among variables		
Not showing causality	False substantive inferences.	Use experiments to confirm causality. Granger's causality test can be used in panel data.
Not accounting for endogeneity	Biased estimates.	Use experiments/panel data to alleviate endogeneity concerns due to reverse causality. Use instrumental variables to account for endogeneity concerns due to an independent variable being a choice variable.
Interaction models		
Mean/residual centring to alleviate multicollinearity	Neither alleviate collinearity. Residual-centring leads to uninterpretable simple effects; mean-centring leads to simple effects that are mathematically equivalent to uncentred models.	Use multiple diagnostics to diagnose collinearity. Also, randomly select and estimate sub-samples to ascertain the stability and plausibility of coefficients. If collinearity is suspected, increase sample sizes to mitigate the loss of power associated with collinearity.
Use of 'main effects only' and 'interaction' models separately	Estimating a 'main effects only' model will lead to an omitted variable bias.	Estimate simple effects and interaction effects simultaneously in a full model.
Omission of 'simple effects' in interaction models	Interaction term becomes uninterpretable.	Estimate simple effects and interaction effects simultaneously in a full model.

Table I. *Continued*

Problem/error	*Consequence*	*Solution*
Structural models		
SEM or PLS?	Under certain conditions, PLS estimates approximate SEM estimates. Otherwise the results may differ.	Use SEM for confirmatory theory-testing research; PLS for exploratory research.
Reliance on goodness of fit for SEM	Good fitting models could occur in the presence of poor loadings. Reliance on fit ignores effect sizes.	Use multiple diagnostics: fit indices, factor loadings, and structural paths to choose the best model.
Small sample size in PLS	Unstable coefficients and large standard errors when estimated with small sample sizes.	Use appropriate sample sizes. Examine and report stability of coefficients and variability in standard errors when using small sample sizes.
Dichotomizing continuous data when testing for moderation effects	Leads to a loss of information and hence reduced power.	Do not dichotomize continuous data. If you must, provide robustness tests to demonstrate that the results do not change across different specifications of the dichotomizing threshold.
Ignoring nested structure of data	Smaller standard errors and incorrect inferences.	Use hierarchical linear models to model nested data.
Longitudinal data		
Confusion between GLMM and GEE	Estimating one model, instead of the other will lead to false substantive inferences.	If the focus is on specific subjects, use mixed models. If the focus is on the population or cohort then use GEE models.
Failure to model error structure in SEM	Increased Type I error rates.	Use longitudinal SEM approaches that incorporate time series processes into an SEM framework when measurement errors are found to correlate across occasions.
Tying results to theory		
Inferences regarding causality	False substantive inferences.	Be cautious when designing studies, and do not over-claim when interpreting results.
Ignoring 'economic' or 'managerial' significance	Statistically significant estimates may be economically insignificant and thus practically meaningless.	Frame and discuss results so that readers understand if the effects are economically significant.
Ignoring importance of marginal effects in non-linear models	Using the sign and significance of the coefficients alone to interpret the model is not appropriate.	Use marginal effects computed as theoretically appropriate to discuss the correct magnitude of an independent variable.

Linkages across Levels of Analysis

There is a growing awareness of the need to incorporate multiple levels of analysis to 'better understand the holistic and interrelated nature of complex organizations' (Rynes, 2005). While construction of such data from scratch can be prohibitively expensive, multilevel linked data are now available through the various government agencies (e.g. US Census, UK Office of National Statistics, Statistics Sweden). Such data are typically longitudinal and near universal (Campbell, 2005), and serve as a strong base for examination of the interaction of industry, firm and individual dynamics. While linked data are popular in the economics literature (Abowd et al., 1999, 2004; Burgess et al., 2000), the rich opportunities provided by the linked datasets for analysis of management issues are somewhat underutilized. The usage of linked datasets presents some access challenges and requires sophisticated statistical techniques, but a few management studies have started tapping these possibilities, particularly for corporate strategy and technology management. For example, Harris et al. (2005) use linked firm-plant data from the UK to examine the effects of corporate restructuring; Siegel and Simons (2006) use Swedish linked employer–employee data to examine the impact of mergers and acquisitions on firm performance, plant productivity and worker outcomes. The size of the datasets provides researchers with opportunities to examine the interactions of firm and individual outcomes with sufficient statistical power.

Combining Qualitative and Quantitative Methodologies

Finally, we would like to note that reliance on quantitative data alone may cause researchers to miss key elements of the phenomena they intend to study (e.g. Lehmann, 2003). Given problems of causal inferences in cross-sectional data, and biases, even in longitudinal data, that result from incomplete inclusion/measurement of all elements in a population, self-reporting and memory recall (Chandy et al., 2004), some researchers advocate the use of the historical approach to data collection (Golder, 2000). This approach encompasses the collection of both qualitative and quantitative data, verification, and interpretation of evidence from multiple sources from the past (Golder, 2000), thereby enabling researchers to accurately reconstruct causal chains of events surrounding a phenomenon (Chandy et al., 2004). In addition to providing insights from the qualitative data as highlighted in Shah and Corley (2006), the gathering of quantitative data (cf. Golder and Tellis, 1993) allows for statistical analysis of data. Further, tapping vast and diverse data sources on a large number of entities helps eliminate self-report biases and creates superior generalizability of conclusions (Chandy et al., 2004).

Shah and Corley (2006) highlight the use of qualitative methodology for theory development, and also provide examples of successful pairings of qualitative and quantitative methods. We endorse their view, and urge researchers that have focused only on quantitative methods to look for ways in which qualitative methods may enhance their research. Hall and Ziedonis (2001), one of the most highly cited articles in Economics and Business in the year after it was published is one such exemplar. While a long line of studies in the economics literature used quantitative estimations of 'patent production functions', the authors' interviews with industry representatives enabled them to identify new variables that were not examined in previous studies.[14] Another example is the study by Klepper and Sleeper (2005), which painstakingly combines historical quantitative and qualitative data to investigate issues related to employee entrepreneurship. Such combined applications of both induction and deduction can help in our goal of addressing unresolved issues in management research.

There is the oft-heard refrain that 'when all you have is a hammer, every problem looks like a nail'. However, not every research problem is a nail, nor can it be solved with just one tool. Poor or inaccurate tests of a theory bring to question its veracity. As a result, it is important for researchers to think creatively about what the best tool or combination of tools for a research question may be so as to address it in an optimal manner. By implementing better practice estimation techniques, extending the levels of analysis, or combining quantitative and qualitative techniques, researchers can advance the frontiers of the field, both by asking new questions and by developing stronger answers to existing questions.

NOTES

*All authors contributed equally, and are listed in reverse alphabetical order. We appreciate comments from Inigo Arroniz, Babu John Mariadoss, Glenn Hoetker, Sonali Shah, and Mike Wright. The usual disclaimer applies.

[1] If only one variable is poorly measured, one can follow the methodology of Fuller and Hidiroglou (1978) (operationalized in Card and Lemieux, 1996) to estimate the reliability of the measures and then rescale the estimates and standard errors for the bias.

[2] An instrumental variable can be viewed as an additional item of the same construct. See Bollen (1996) on the use of 2SLS to estimate both measurement and structural models in SEM.

[3] We note that the original Heckman sample selection correction, a special case of 2SLS, is currently under debate particularly due to its sensitivity to distributional assumptions (see Greene (1997) for more details).

[4] This critical value is for the common case of one instrumental variable in lieu of one endogenous variable. At this critical value, a 5 per cent hypothesis test of the second stage estimates rejects less than 15 per cent of the time.

[5] Simultaneous causality can be viewed as an omitted variable problem.

[6] See Larcker and Rusticus (2005) for an overview of instrumental variables validity tests as well as cautions on the appropriateness of using instrumental variables techniques to address endogeneity issues.

[7] The inability to use laboratory experiments may arise due to either cost or ethical considerations (e.g. in instances where denying a treatment is unethical).

[8] Echambadi and Hess (2007) show that the magnitude of the covariance between x_1 and x_1x_2 can sometimes increase with mean-centring of non-symmetric random variables.

[9] Standardized loadings of 0.7 imply a shared variance of 49 per cent, or approximately half the item variance is explained by the construct.

[10] Typical longitudinal data that involve a set of repeated observations over time on a group of units is inherently hierarchical and can be modelled using HLM approaches.

[11] Recent approaches in structural equation modelling have attempted to specify time series processes for longitudinal data in an SEM framework (see Sivo et al. (2006) for more details).

[12] The marginal effect of an independent variable is the first derivative of the conditional mean of the dependent variable with respect to that variable (Bowen and Wiersema, 2005). In the case of non-linear models, owing to the non-linear nature of the conditional mean, this cross-derivative must be derived analytically (Ai and Norton, 2003; Bowen and Wiersema, 2005) or by using bootstrapping techniques (Echambadi et al., 2006b).

[13] The marginal effect at the average can sometimes be problematic because there may be no meaningful 'average' observation (e.g. when gender is a relevant explanatory variable, the marginal effect at the average captures a non-existent statistic).

[14] ISI interview of Rosemarie Ziedonis, available at http://www.esi-topics.com/nhp/comments/november-02-RosemarieZiedonis.html.

REFERENCES

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). 'High wage workers and high wage firms'. *Econometrica*, **67**, 251–333.

Abowd, J. M., Haltiwanger, J. and Lane, J. (2004). 'Integrated longitudinal employer-employee data for the United States'. *The American Economic Review*, **94**, 224–9.

Agarwal, R., Sarkar, M. B. and Echambadi, R. (2002). 'The conditioning effect of time on firm survival: an industry life cycle approach'. *Academy of Management Journal*, **45**, 971–94.

Ai, C. and Norton, E. C. (2003). 'Interaction terms in logit and probit models'. *Economics Letters*, **80**, 123–9.

Aiken, L. S. and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Allison, P. D. (2005). *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Institute.

Amemiya, T. (1985). *Advanced Econometrics*. Boston, MA: Harvard University Press.

Ashenfelter, O. and Card, D. (1985). 'Using the longitudinal structure of earnings to estimate the effect of training programs'. *Review of Economics and Statistics*, **67**, 648–60.

Ballinger, G. A. (2004). 'Using generalized estimating equations for longitudinal data analysis'. *Organizational Research Methods*, **7**, 127–50.

Barclay, D., Higgins, C. and Thompson, R. (1995). 'The partial least squares (PLS) approach to causal modeling: personal computer adoption and use as an illustration'. *Technology Studies*, **2**, 285–324.

Berg, G. D. and Mansley E. C. (2004). 'Endogeneity bias in the absence of unobserved heterogeneity'. *Annals of Epidemiology*, **14**, 561–5.

Blalock, H. M. (1964). *Causal Inferences in Nonexperimental Research*. Chapel Hill, NC: University of North Carolina Press.

Boland, R. J. Jr, Singh, J., Salipante, P., Aram, J. D., Fay, S. Y. and Kanawattanachai, P. (2001). 'Knowledge representations and knowledge transfer'. *Academy of Management Journal*, **44**, 393–417.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, K. A. (1996) 'An alternative two stage least squares (2SLS) estimator for latent variable equations'. *Psychometrika*, **61**, 109–21.

Bowen, H. and Wiersema, M. (2005). 'Foreign-based competition and corporate diversification strategy'. *Strategic Management Journal*, **26**, 1153–71.

Boyd, B. K., Gove, S. and Hitt, M. A. (2005a). 'Consequences of measurement problems in strategic management research: the case of Amihud and Lev'. *Strategic Management Journal*, **26**, 367–75.

Boyd, B. K., Gove, S. and Hitt, M. A. (2005b). 'Construct measurement in strategic management research: illusion or reality?' *Strategic Management Journal*, **26**, 239–57.

Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Burgess, S., Lane, J. and Stevens, D. (2000). 'Job flows, worker flows, and churning'. *Journal of Labor Economics*, **18**, 473–502.

Campbell, B. A. (2005). 'Using linked employer-employee data to study entrepreneurial issues'. In Alvarez, S. A., Agarwal, R. and Sorenson, O. (Eds), *Handbook of Entrepreneurship Research: Disciplinary Perspectives*. New York: Springer.

Campbell, B. A. (2006). 'Entrepreneurial Risk and Return'. Working Paper. Center for Human Resources, The Wharton School, University of Pennsylvania.

Card, D. and Lemieux, T. (1996). 'Wage dispersion, returns to skill, and black-white wage differentials'. *Journal of Econometrics*, **74**, 319–61.

Chandy, R., Golder, P. and Tellis, G. (2004). 'Historical research in marketing strategy: method, myths, and promise'. In Moorman, C. and Lehmann, D. (Eds), *Cools Tools for Assessing Marketing Strategy Performance*. Boston, MA: Marketing Science Institute.

Chay, K. Y., McEwan, P. J. and Urquiola, M. (2005). 'The central role of noise in evaluating interventions that use test scores to rank schools'. *American Economic Review*, **95**, 1237–58.

Chin, W. W. (1998). 'Issues and opinion on structural equation modeling'. *MIS Quarterly*, **22**, 7–16.

Chin, W. W. and Newsted, P. R. (1999). 'Structural equation modeling analysis with small samples using partial least squares'. In Hoyle, R. (Ed.), *Statistical Strategies for Small Sample Research*. Thousand Oaks, CA: Sage, 307–41.

Chin, W. W., Marcolin, B. L. and Newsted, P. R. (2003). 'A partial least squares latent variable modeling approach for measuring interaction effects: results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study'. *Information Systems Research*, **14**, 189–217.

Christensen, C. M. (2006). 'The ongoing process of building a theory of disruption'. *The Journal of Product Innovation Management*, **23**, 39–55.
Churchill, G. A. Jr (1979). 'A paradigm for developing better measures of marketing constructs'. *Journal of Marketing Research*, **16**, 64–73.
Cohen, J. (1990). 'Things I have learned (so far)'. *American Psychologist*, **12**, 1304–12.
Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings*. Boston, MA: Houghton Mifflin Company.
Cortina, J. M., Chen, G. and Dunlap, W. P. (2001). 'Testing interaction effects in LISREL: examination and illustration of available procedures'. *Organizational Research Methods*, **4**, 324–60.
Dehejia, R. and Wahba, S. (2002). 'Propensity score matching methods for non-experimental causal studies'. *Review of Economics and Statistics*, **84**, 151–61.
Dekimpe, M. G. and Hanssens, D. M. (1995). 'Empirical generalizations about market evolution and stationarity'. *Marketing Science*, **14**, G109–21.
Dekimpe, M. G. and Hanssens, D. M. (1999). 'Sustained spending and persistent response: a new look at long-term marketing profitability'. *Journal of Marketing Research*, **36**, 397–412.
Diamantopoulos, A. and Winklhofer, H. M. (2001). 'Index construction with formative indicators: an alternative to scale development'. *Journal of Marketing Research*, **38**, 269–77.
Echambadi, R. and Hess, J. D. (2007). 'Mean-centering does not alleviate collinearity problems in moderated multiple regression models'. *Marketing Science*, forthcoming.
Echambadi, R., Arroniz, I., Reinartz, W. and Lee, J. (2006a). 'Empirical generalizations from brand extension research: how sure are we?'. *International Journal of Research in Marketing*, **23**, 253–61.
Echambadi, R., Sarkar, M. B., Agarwal, R. and Arroniz, I. (2006b). 'Entry Timing and Firm Survival: The Moderating Effects of Firm Size and Pace of Technological Change'. University of Central Florida, Working Paper.
Farrell, A. D. (1994). 'Structural equation modeling with longitudinal data: strategies for examining group differences and reciprocal relationships'. *Journal of Consulting and Clinical Psychology*, **62**, 477–87.
Fornell, C. and Larcker, D. F. (1981). 'Evaluating structural equation models with unobservable variables and measurement error'. *Journal of Marketing Research*, **18**, 39–50.
Fuller, W. A. and Hidiroglou, M. A. (1978). 'Regression estimation after correcting for attenuation'. *Journal of the American Statistical Association*, **73**, 99–104.
Goldberger, A. (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.
Golder, P. N. (2000). 'Historical method in marketing research with new evidence on long-term market share stability'. *Journal of Marketing Research*, **37**, 156–72.
Golder, P. N. and Tellis, G. J. (1993). 'Pioneer advantage: marketing logic or marketing legend?'. *Journal of Marketing Research*, **30**, 158–70.
Granger, C. W. J. (1969). 'Investigating causal relations by econometric models and cross-spectral methods'. *Econometrica*, **37**, 424–38.
Granger, C. W. J. (2003). 'Evaluation of theories and models'. In Stigum, B. (Ed.), *Econometrics and the Philosophy of Economics*. Princeton, NJ: Princeton University Press.
Greene, W. H. (1997). *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.
Hall, B. H. and Ziedonis, R. H. (2001). 'The patent paradox revisited: an empirical study of patenting in the U.S. semiconductor industry, 1979–1995'. *The Rand Journal of Economics*, **32**, 101–28.
Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
Harris, R., Siegel, D. and Wright, M. (2005). 'Assessing the impact of management buyouts on economic efficiency: plant-level evidence from the United Kingdom'. *The Review of Economics and Statistics*, **87**, 148–53.
Hausman, J. (1978). 'Specification tests in econometrics'. *Econometrica*, **46**, 1251–72.
Heckman, J. (1979). 'Sample selection bias as a specification error'. *Econometrica*, **47**, 153–61.
Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*. New York: John Wiley and Sons.
Hoetker, G. (2007). 'The use of logit and probit models in strategic management research: critical issues'. *Strategic Management Journal*, forthcoming.
Hofmann, D. A. (1997). 'An overview of the logic and rationale of hierarchical linear models'. *Journal of Management*, **23**, 723–44.
Hofmann, D. A. and Gavin, M. B. (1998). 'Centering decisions in hierarchical linear models: implications for research in organizations'. *Journal of Management*, **24**, 623–41.
Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R. and Pentz, M. A. (1998). 'Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes'. *American Journal of Epidemiology*, **147**, 694–703.

Hulland, J. (1999). 'Use of partial least squares (PLS) in strategic management research: a review of four recent studies'. *Strategic Management Journal*, **20**, 195–204.

Irwin, J. R. and McClelland, G. H. (2001). 'Misleading heuristics and moderated multiple regression models'. *Journal of Marketing Research*, **38**, 100–9.

Irwin, J. R. and McClelland, G. H. (2003). 'Negative consequences of dichotomizing continuous predictor variables'. *Journal of Marketing Research*, **40**, 366–71.

Jarvis, C. B., MacKenzie, S. B. and Podsakoff, P. M. (2003). 'A critical review of construct indicators and measurement model misspecification in marketing and consumer research'. *Journal of Consumer Research*, **30**, 199–218.

Kerlinger, F. N. (1973). *Foundations of Behavioral Research*, 2nd edition. New York: Holt, Rinehart and Winston.

Klepper, S. and Sleeper, S. (2005) 'Entry by spinoffs'. *Management Science*, **51**, 1291–2007.

Lance, C. E. (1988). 'Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions'. *Applied Psychological Measurement*, **12**, 163–75.

Larcker, D. and Rusticus, T. (2005). 'On the Use of Instrumental Variables in Accounting Research'. Working Paper, Graduate School of Business, Stanford University.

Lehmann, D. (2003). *The Relevance of Rigor*. Report No. 03-105. Cambridge, MA: Marketing Science Institute.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences*. Thousand Oaks, CA: Sage.

MacCallum, R. C. and Browne, M. W. (1993). 'The use of causal indicators in covariance structure models: some practical issues'. *Psychological Bulletin*, **114**, 533–41.

MacKenzie, S. B. (2001). 'Opportunities for improving consumer research through latent variable structural equation modeling'. *Journal of Consumer Research*, **28**, 159–66.

MacKenzie, S. B. (2003). 'The dangers of poor construct conceptualization'. *Academy of Marketing Science Journal*, **31**, 323–6.

Marcoulides, G. A. and Saunders, C. (2006). 'PLS: a silver bullet'. *MIS Quarterly*, **30**, 1–7.

Mason, C. H. and Perreault, W. D. Jr (1991). 'Collinearity, power, and interpretation of multiple regression analysis'. *Journal of Marketing Research*, **28**, 268–80.

Meehl, P. E. (1990). 'Why summaries of research on psychological theories are often uninterpretable'. *Psychological Reports*, **66**, 195–244.

Mela, C. F. and Kopalle, P. (2002). 'The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations'. *Applied Economics*, **34**, 657–77.

Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablynski, C. J. and Erez, M. (2001). 'Why people stay: using job embeddedness to predict voluntary turnover'. *Academy of Management Journal*, **44**, 1102–21.

Nunnally, J. C. and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd edition. New York: McGraw-Hill.

Ofir, C. and Khuri, A. (1986). 'Multicollinearity in marketing models: diagnostics and remedial measures'. *International Journal of Research in Marketing*, **3**, 181–205.

Rynes, S. (2005). 'Taking stock and looking ahead'. *Academy of Management Journal*, **48**, 9–15.

Sarkar, M. B., Echambadi, R. and Harrison, J. S. (2001). 'Alliance entrepreneurship and firm market performance'. *Strategic Management Journal*, **22**, 701–11.

Shah, S. and Corley, K. (2006). 'Building better theory by bridging the quantitative-qualitative divide'. *Journal of Management Studies*, **43**, 1823–37.

Shook, C. L., Ketchen, D. J. Jr, Hult, G. T. M. and Kacmar, K. M. (2004). 'An assessment of the use of structural equation modeling in strategic management research'. *Strategic Management Journal*, **25**, 397–404.

Siegel, D. and Simons, K. (2006). 'Assessing the Effects of Mergers and Acquisitions on Firm Performance, Plant Productivity and Workers: New Evidence from Matched Employer-Employee Data'. Rensselaer Working Paper in Economics #0601.

Sivo, S. A., Fan, X. and Witta, L. (2005). 'The biasing effects of unmodeled ARMA time series processes growth curve model estimates'. *Structural Equation Modeling: A Multidisciplinary Journal*, **12**, 215–31.

Sivo, S. A., Echambadi, R. and Arroniz, I. (2006). 'Incorporating Time Series Processes in Structural Equation Models'. Working Paper, University of Central Florida.

Spector, P. E. (1997). *Job Satisfaction*. Thousand Oaks, CA: Sage.

Stock, J. and Yogo, M. (2002). 'Testing for weak instruments in linear IV regression'. In Andrews, D. W. K. and Stock, J. (Eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 80–108.

Stock J., Wright, J. and Yogo, M. (2002). 'A survey of weak instruments and weak identification in generalized method of moments'. *Journal of Business and Economic Statistics*, **20**, 518–29.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.
Zeger, S. L., Liang, K. Y. and Albert, P. A. (1988). 'Models for longitudinal data: a generalized estimating equation approach'. *Biometrics*, **44**, 1049–60.
Zohoori, N. and Savitz, D. A. (1997). 'Econometric approaches to epidemiologic data: relating endogeneity and unobserved heterogeneity to confounding'. *Annals of Epidemiology*, **7**, 251–7.