The University of Southern Mississippi
The Aquila Digital Community

Faculty Publications

Using Partial Least Squares in Operations Management Research: A Practical Guideline and Summary of Past Research

David Xiaosong Peng
Texas A&M University, xpeng@bauer.uh.edu

Fujun J. Lai
University of Southern Mississippi, fujun.lai@usm.edu

Follow this and additional works at: https://aquila.usm.edu/fac_pubs

Part of the Business Commons

Using Partial Least Squares in Operations Management Research: A Practical Guideline and Summary of Past Research

David Xiaosong Peng

Department of Information & Operations Management
Mays Business School at Texas A&M University
320 Wehner Building | 4217 TAMU
College Station, TX 77843-4217
Phone: 979.845.6996; Fax: 979.845.5653; E-mail: xpeng@mays.tamu.edu

*Fujun Lai

Department of Management and International Business
College of Business
University of Southern Mississippi
730 E. Beach Blvd,
Long Beach, MS 39503
Phone: 228.214.3446; Fax: 228.214.3475; E-mail: Fujun.Lai@usm.edu

*Corresponding author

*Response to Reviewers

Dear dpeng,

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript, no later than 240 days from today.

To submit a revision, please go to http://ees.elsevier.com/opeman/ and login as an Author.
Your username is: dpeng
If you need to retrieve password details, please go to:
http://ees.elsevier.com/opeman/automail_query.asp

On your Main Menu page is a folder entitled "Submissions Needing Revision". You will find your submission record there.

Yours sincerely,

Thomas Younghoon Choi, PhD
Editor-in-Chief
Journal of Operations Management

Reviewers' comments:

Reviewer #4: I want to thank the authors for their response to the comments the reviewers and I made. I think that they have done a good job with the revision. I have several issues that I would like to be addressed.

> **Author response**: Thank you. We are encouraged by your above comments. We believe we have addressed your remaining concerns through this round of revision. Please see the revised manuscript and our itemized responses to your comments below.

1. In the third paragraph of Section 3, the paper mentions that intra-construct item correlations are greater than the inter-construct item correlation. The paper also mentions in this same paragraph that the item-to-own-construct correlations are greater than item-to-other-construct correlations. I would like to ask the authors to provide statistical test support for these clams. Are these differences statistically significant?

> **Author response**: We performed statistical tests for the above comparison. As an example, we examined the item correlations within the *operational performance* (OP) construct and between this construct and the construct *trust with suppliers (TRUST).*
>
> First, we compared the intra-construct item correlations within OP (six correlations) and the inter-construct item correlations between OP and TRUST (12 correlations). We found that the intra-construct item correlations are clearly larger than the inter-construct item correlations. We also compared the average of the above mentioned six intra-construct item correlations and the average of the 12 inter-construct item correlations. We found that the average of the intra-construct item correlations is significant larger than the average of the inter-construct item correlations (p<0.01).
>
> Second, we computed OP's item-to-own-construct correlations (four correlations) and its item-to-other-construct correlations (here the other construct is *Trust*, there are four correlations). Next, we compared the average of the item-to-own-construct correlations and the average of the item-to-other-construct correlations. We found the difference statistically significant (p<0.01).
>
> We revisited the original study (Klein and Rai, 2009) that proposes this method of examining discriminant validity of formative constructs. The study simply compares the magnitude of item-to-own-construct correlations and item-to-other-construct correlations. But the study does not perform a formal statistical test of the above comparison. Our conjecture is that the authors only visually checked the magnitude of those correlations. Thus, we revised our paper and point out that this method of examining formative construct discriminant validity is quite new and there is no detailed guidance on applying this method. For this reason, we decide to deemphasize this method in our guideline (Section 2) and example (Sections 3).

2. In the fifth paragraph in Section 3, you claim that the results are consistent. But you don't clarify what they are consistent in regards to. Please expand on this.

> **Author response**: We now state "the magnitude and significance of structural paths are significant".

3. At the start of Section 4, please expand on the first sentence by providing a one-two sentence summary of why you are reviewing the literature. The first sentence in this section reads: "This section reviews PLS use in the recent OM articles." Why are you reviewing the literature? In order to provide what in particular?

> **Author response**: We expanded the sentence as follows:
>
> "This section reviews PLS use in recent OM literature. This review allows us to identity which aspects of PLS researchers should pay attention to and also serves as the starting point for creating our guideline for evaluating and using PLS."

4. In the fifth paragraph in Section 2.3.2, the paper states that "while the "10-times" heuristic for determining the sample size adequacy in using PLS has been widely adopted, it has been criticized for not taking into account the effect size, reliability, number of indicators, or other factors which in one way or another may affect statistical power of a hypothesis test (Goodhue et al., 2006; Marcoulides and Sanders, 2006)." I am sorry but this sentence makes no sense the heuristic you referred to here tries to address the issues you claim is being criticized for. I think you should expand on this and write this sentence more thoroughly

> **Author response**: We agree the above sentence is not very clear. We have modified and expanded the sentence as follows:
>
> "The effect size, reliability, the number of indicators, or other factors may affect the statistical power of a hypothesis test. Simply applying the "10 times" rule of thumb may lead researchers to underestimate the sample size requirement in certain situations, such as small effect sizes and low reliability of measurement items. In other words, applying the "10 times" rule of thumb without performing a formal power analysis may lead to hypothesis tests with low power."

5. The paper is fairly well written. However, I think it would be in the best interest of the authors to have it copy edited by a professional. I found a number of problem areas. Please see below:

> **Author response**: We used a professional to copy edit our paper. The authors of the paper also carefully edited the paper. We hope our efforts have improved the paper to your satisfaction.

In the third paragraph of Secion 1: "Empirical OM researchers face some unique challenges such as the relatively less developed empirical knowledge (Wacker, 1998), the lack of standardized measurement scales (Roth et al., 2007), and the difficulty to obtain large samples since OM researchers typically examine phenomenon at the firm level or the supply chain level."

THIS SENTENCE SHOULD READ

"Empirical OM researchers face some unique challenges such as the relatively less developed empirical knowledge (Wacker, 1998), the lack of standardized measurement scales (Roth et al., 2007), and the difficulty to obtain large samples since OM researchers typically examine phenomena at the firm level or the supply chain level."

Author response: changed as suggested.

In the second paragraph of Section 2.2.1: "If a researcher wants to evaluate if the research model still holds today, from a theoretical standpoint, the study should be confirmatory in nature since the theory underlying the research model to be tested is well established."

THIS SENTENCE SHOULD READ.

"To evaluate if the research model still holds today from a theoretical standpoint, a study should be confirmatory in nature since the theory underlying the research model to be tested is well established."

Author response: changed as suggested.

In the third paragraph of Secion 2.2.4: "As part of the efforts to establish the content validity of formative constructs, researchers are recommended to conduct a thorough literature review related to the construct domain."

THIS SENTENCE SHOULD READ.

"As part of the efforts to establish the content validity of formative constructs, we recommend that researchers conduct a thorough literature review related to the construct domain."

Author response: changed as suggested.

In the last two paragraphs of Section 2.2:

"For instance, conceptually, an item measuring flexibility does not have to highly correlate with an item measuring manufacturing costs."

THIS SENTENCE SHOULD READ.

"For instance, conceptually, an item measuring flexibility does not have to correlate with an item measuring manufacturing costs."

Author response: changed as suggested.

"Fourth, with respect to nomological network, one cannot expect that different operational performance items are impacted by the same set of antecedents and lead to the same set of consequences."

THIS SENTENCE SHOULD READ.

"Fourth, with respect to nomological networks, one cannot expect that different operational performance items will be impacted by the same set of antecedents and lead to the same set of consequences."

Author response: changed as suggested.

"Since a formative construct by itself is under-identified, researcher should consider including two or more reflective indicators for each formative construct in their survey."

THIS SENTENCE SHOULD READ.

"Since a formative construct by itself is under-identified, researchers should consider including two or more reflective indicators for each formative construct in their measurements."

Author response: changed as suggested.

"These reflect indicators usually are not a part of the research model to be tested but rather are used as an external criterion to assess the formative construct validity (Diamantopoulos and Winklhofer, 2001)."

THIS SENTENCE SHOULD READ.

"These reflective indicators are not usually a part of the research model to be tested but rather are used as an external criterion to assess the formative construct validity (Diamantopoulos and Winklhofer, 2001)."

Author response: changed as suggested.

Sixth paragraph in Section 2.3.1.: "Third, researchers should check external validity of formative constructs."

THIS SENTENCE SHOULD READ.

"Third, researchers should check the external validity of formative constructs."

Author response: changed as suggested.

Sixth paragraph in Section 2.3.2: "Fourth, the post-hoc power analysis should be conducted to check if the power of the research study is acceptable (>0.80)."

THIS SENTENCE SHOULD READ.

"Fourth, post-hoc power analyses should be conducted to check if the power of the research study is acceptable (>0.80)."

Author response: changed as suggested.

Seventh paragraph in Section 2.3.2: "Fifth, although PLS does not provide overall fit statistics, recently researchers have begun to compute Goodness of Fit (GoF) in PLS (Tenenhaus et al., 2005), which

considers the quality of the complete measurement model in terms of average communality (i.e., AVE) and the quality of the complete structural model in terms of average R^2."

THIS SENTENCE SHOULD READ.

"Fifth, although PLS does not provide overall fit statistics, researchers have recently begun to compute Goodness of Fit (GoF) in PLS (Tenenhaus et al., 2005), which considers the quality of the complete measurement model in terms of average communality (i.e., AVE) and the quality of the complete structural model in terms of average R^2. "

      Author response: changed as suggested.

Third paragraph in Section 3: "For each formative item, we examine the item's item weight (rather than item loadings), sign, and magnitude."

THIS SENTENCE SHOULD READ.

"For each formative item, we examine its weight (rather than item loadings), sign, and magnitude."

      Author response: changed as suggested.

Fifth paragraph in Section 3: "We run the structural model using the bootstrap procedure with 200, 500, and 1,000 times of resampling."

THIS SENTENCE SHOULD READ.

"We ran the structural model using the bootstrap procedure with 200, 500, and 1,000 times of resampling."

      Author response: changed as suggested. We appreciate your meticulous review.

      Author response: Please note that we have edited the paper according to your above suggestions. However, through the copy editing by a professional and multiple rounds of copy editing by the authors after accepting your suggested changes, in some instances the final version of the paper may be slightly different from the change you suggested.

Using Partial Least Squares in Operations Management Research: A Practical Guideline and Summary of Past Research

ABSTRACT

The partial least squares (PLS) approach to structural equation modeling (SEM) has been widely adopted in business research fields such as information systems, consumer behavior, and marketing. The use of PLS in the field of operations management is also growing. However, questions still exist among some operations management researchers regarding whether and how PLS should be used. To address these questions, our study provides a practical guideline for using PLS and uses examples from the operations management literature to demonstrate how the specific points in this guideline can be applied. In addition, our study reviews and summarizes the use of PLS in the recent operations management literature according to our guideline. The main contribution of this study is to present a practical guideline for evaluating and using PLS that is tailored to the operations management field.

Keywords: Partial least squares (PLS); Structural equation modeling (SEM); Empirical research methods; Operations management

Using Partial Least Squares in Operations Management Research: A Practical Guideline and Summary of Past Research

1. INTRODUCTION

Structural equation modeling (SEM) has been widely adopted in social and psychological research. Operations management (OM) researchers have also used SEM to a great extent (Shah and Goldstein, 2006). To date, OM researchers have mainly adopted covariance-based SEM (CBSEM) methods, as exemplified by software such as LISREL, AMOS, and EQS. A less widespread technique known as partial least squares (PLS) has started to receive attention from OM researchers, as evidenced by the steady growth of PLS use in the OM field.

As an SEM method, PLS has been subjected to much debate with respect to its pros and cons and under what circumstances it should be adopted, if at all. Advocates of PLS claim that it has the ability to estimate research models using small samples with no strict distribution assumptions and can model both reflective and formative constructs within the same research model. PLS also supposedly avoids the inadmissible solutions and factor indeterminacy of CBSEM (Chin, 1998b). Researchers who oppose using PLS cite reasons such as bias in parameter estimates, its inability to model measurement errors, and its piecemeal approach to estimating the overall research model.

Despite the controversies and debate surrounding PLS, interest in PLS among OM researchers seems to be growing. Although a number of articles and book chapters have summarized PLS algorithms, reviewed the use of PLS in a research field, or discussed specific aspects of PLS applications such as sample size requirements and specifying formative constructs, we are not aware of any guideline for evaluating and using PLS that is tailored to the

OM audience. Empirical OM researchers face some unique challenges such as relatively less developed empirical knowledge (Wacker, 1998), a lack of standardized measurement scales (Roth et al., 2007), and the difficulty of obtaining large samples because OM researchers typically examine phenomena at the firm or the supply chain level. These challenges may limit the applicability of CBSEM. Consequently, OM researchers should evaluate different analysis techniques, particularly PLS if SEM is preferred. To help OM researchers evaluate and use PLS, this study provides a practical guideline that outlines some of the important issues in using PLS. We make this guideline specific to the OM field by using illustrative examples from the OM literature.

We also summarize studies that use PLS to examine OM topics in the fields of operations management, strategic management, and organization theory from 2000 to 2011. We review these articles with respect to their rationales for using PLS, sample sizes, the use and assessment of formative constructs, bootstrapping procedures, and the presentation of results. Our review provides a mixed picture of PLS use in the OM field, with some studies exhibiting deficiencies or lack of familiarity with certain aspects of PLS and others demonstrating a reasonably good understanding of the PLS method.

To the best of our knowledge, this study is the first to provide a practical guideline for using PLS that includes illustrative examples from the OM literature. This guideline can serve as a useful checklist for OM researchers in their evaluations regarding whether PLS can meet their data analysis needs given their research objectives, research model characteristics, sample sizes, and sample distribution. In addition, our study performs a thorough review of the use of PLS in

the OM literature. This review highlights the common problems of using PLS and thus can help OM researchers avoid similar mistakes in future studies.

2. A GUIDELINE FOR EVALUATING AND USING PLS

2.1. PLS Overview

PLS, originally introduced by Wold in the 1960s (Wold, 1966), was recently revitalized by Chin in the information systems (IS) field (Chin, 1998a, b; Chin et al., 2003). In addition to OM, PLS has been used in management (e.g., Cording et al., 2008), marketing (e.g., Hennig-Thurau et al., 2006; White et al., 2003), strategic management (Hulland, 1999), and other business research fields. Representative PLS software tools include *PLS-Graph* and *SmartPLS,* among others. Appendix 1 provides a non-technical introduction to the PLS algorithm used by the most popular PLS software: *PLS-Graph*. In-depth coverage of this PLS algorithm can be found in Chin and Newsted (1999).

One major difference between CBSEM and PLS is that the former focuses on common factor variances and the latter considers both common and unique variances (i.e., overall variances). The difference between CBSEM and PLS is similar to that between common factor analysis and principle component analysis (Chin, 1995). CBSEM specifies the residual structure of latent variables, whereas in PLS, the latent variables are weighted composite scores of the indicator variables and lead directly to explicit factor scores.

PLS is also less well grounded in statistical theory than CBSEM to the extent that it is considered statistically inferior (Chin, 1995). PLS estimators do not have the precision of maximum likelihood (ML) estimation (as used in CBSEM, such as LISREL) in achieving optimal predictions. When the multivariate normality assumption is met, CBSEM estimates are

efficient in large samples and support analytical estimates of asymptotic standard errors. In contrast, because the construct scores of the latent variables in PLS are created by aggregating indicator items that involve measurement errors, PLS estimates of construct scores are biased and are only consistent under the conditions of "consistency at large", which refer to a large number of items per construct, high communality, and large sample sizes (Wold, 1982, p.25). Because PLS lacks a classical parametric inferential framework, parameters are estimated using resampling procedures such as *bootstrap* and *jackknife*.

We suggest that OM researchers use CBSEM if its assumptions are met. However, when the conditions for using CBSEM are not met, researchers should evaluate the pros and cons of CBSEM and PLS and should only use PLS if doing so proves more appropriate overall. We summarize our guideline for evaluating and using PLS in Table 1 and discuss its specific points in detail in the rest of Section 2.

-- Insert Table 1 about here --

2.2. Issues to Consider During the Pre-Analysis Stage

Considerations of construct formulation and analysis techniques should begin in the research design stage. To choose between CBSEM and PLS, researchers should carefully consider the objectives of their study, the state of the existing knowledge about the research model to be tested, the characteristics of the research model (i.e., is the research model extremely complex?), and the conceptualization and formulation of the constructs (i.e., are constructs formative or reflective?).

2.2.1. Research Objectives (Confirmatory versus Exploratory Studies)

PLS aims to assess the extent to which one part of the research model predicts values in other parts of the research model. In this sense, PLS is prediction-oriented (Fornell and Bookstein, 1982; Vinzi et al., 2010). In contrast, CBSEM estimates the complete research model and produces fit statistics that explain how well the empirical data fits the theoretical model (i.e., minimizing the discrepancy between the covariances of sample data and those specified by the theoretical model). As such, CBSEM is parameter-oriented because it seeks to create parameter estimates that are close to population parameters. This difference suggests that CBSEM is more appropriate when there are well-established theories underlying the proposed research model. In such a circumstance, researchers can use CBSEM to obtain population parameter estimates that explain covariances with the assumption that the underlying model is correct. However, if the overall nomological network has not been well understood and researchers are trying to explore relationships among the theoretical constructs and to assess the predictive validity of the exogenous variables, then PLS can be considered.

An illustrative research model that can be tested using CBSEM is the theory of quality management underlying the Deming management method, as described in Anderson et al. (1994). The main tenets of Deming's management methods are well accepted by both scholars and practitioners. Anderson et al. (1994) articulate the theoretical relationships among the constructs in the research model based on the relevant literature, an observation of industry practices, and the results of a Delphi study that assembled a panel of industry and academic experts in quality management. Their research model has since been subjected to empirical validation (Anderson et al., 1995). To evaluate whether their research model still holds from a theoretical standpoint, a study should be confirmatory in nature because the theory underlying the research model to be

tested is well-established. Thus, a main objective of the data analysis should be to find out how well the data collected from the current business environment fit the research model. CBSEM would be appropriate for this end, assuming that the other requirements for CBSEM (e.g., sample sizes and sample distribution) are met.

An example of when PLS might be more appropriate for testing a research model can be found in Cheung et al. (2011). The objective of their study is to explore the extent to which relational learning is associated with the relational performance of both the buyer and the supplier in a supply chain dyad. These relationships had seldom been examined in the literature at the time, and there was no well-established theory that could directly serve as the theoretical foundation of their hypothesized relationships. As such, a main objective of the analysis should be to identify the predictive power of the exogenous variables (a list of proposed drivers of relational performance) on the endogenous variables (relational performance), making PLS a potentially appropriate analysis tool.

2.2.2. Sample Sizes and Model Complexity

Sample Sizes

Sample sizes are an important consideration in SEM because it can affect the reliability of parameter estimates, model fit, and the statistical power of SEM (Shah and Goldstein, 2005). The literature proposes different sample size requirements for CBSEM and PLS. Common sample size rules of thumb for CBSEM suggest examining the ratio of the sample size to the total number of parameters estimated, whereas sample size rules of thumb for PLS usually only suggest examining the ratio of the sample size to the most complex relationship in the research model.

Commonly used rules of thumb for determining sample size adequacy in CBSEM include "establishing a minimum (e.g., 200), having a certain number of observations per measurement item, having a certain number of observations per parameters estimated (Bentler and Chou, 1987; Bollen, 1989), and through conducting [a] power analysis (MacCallum et al., 1996)" (Shah and Goldstein, 2006, p.154). With respect to PLS, the literature frequently uses the "10 times" rule of thumb as the guide for estimating the minimum sample size requirement. This rule of thumb suggests that PLS only requires a sample size of 10 times the most complex relationship within the research model. The most complex relationship is the larger value between (1) the construct with the largest number of formative indicators if there are formative constructs in the research model (i.e., largest measurement equation (LME)) and (2) the dependent latent variable (LV) with the largest number of independent LVs influencing it (i.e., the largest structural equation (LSE)). Researchers have suggested that the "10 times" rule of thumb for determining sample size adequacy in PLS analyses only applies when certain conditions, such as strong effect sizes and high reliability of measurement items, are met. Thus, the literature calls for researchers to calculate statistical power to determine sample size adequacy (Marcoulides and Saunders, 2006).

We use the theoretical framework underlying Deming's management theory (Anderson et al., 1995) as an illustrative example to explain the "10 times" rule of thumb for evaluating sample size adequacy when using PLS. We are not suggesting that PLS is more appropriate for testing the above theoretical model. Because the research model includes only reflective constructs, the most complex relationship is the dependent LV with the largest number of independent LVs influencing it, which would be 2 in this research model. Thus, the minimum sample size requirement can be as low as 20 ($10 \times 2 = 20$) when PLS is used to test the research

model, assuming certain conditions are met (e.g., adequate effect sizes, a sufficiently large number of items per construct, and highly reliable constructs). However, if we follow the rules of thumb for CBSEM sample size requirements, which typically range from 5 (Tanaka, 1987) to 20 (Bentler and Chou, 1987) times the number of parameters estimated, the sample size requirement for testing the same model using CBSEM would be 370 to 1,480 observations (the number of parameters estimated is 74 in the research model, such that $74 \times 5 = 370$ and $74 \times 20 = 1,480$). We note that the above methods for determining sample size requirements are rules of thumb that researchers can use in the pre-analysis stage to make a rough estimate. Researchers still should perform a power analysis to formally determine whether the sample size is adequate for using PLS or CBSEM.

A point related to the sample size issue is the questionnaire design. Because increasing the number of indicators per construct is one way to reduce the bias in the parameter estimate for reflective constructs in PLS, researchers can consider including a large number of items for reflective constructs in the survey questionnaire if they anticipate that PLS may be used in the analysis stage. It should be noted that researchers often face a tradeoff between response rate and questionnaire length, and that increasing the number of items per construct can adversely affect a survey's response rate. Nevertheless, we suggest that researchers take the number of items per construct into consideration during the research design stage.

Model Complexity

The overall complexity of the research model has a direct impact on sample size adequacy in CBSEM, but not necessarily in PLS. Considerations such as multi-level analyses, multiple endogeneity, mediation analyses, moderation analyses, and higher-order factors can

increase the total number of parameter estimates, possibly leading to model identification and convergence issues in CBSEM. For instance, in a multi-level analysis where group size is small and intra-cluster correlation is low, the between-group part of the model may yield an inadmissible solution in CBSEM (Hox and Mass, 2001). A moderation effect in SEM is typically tested via a new construct that uses indicators computed by cross-multiplying the standardized items of each construct involved in the moderation effect (Chin et al., 2003). This cross-multiplying can potentially generate a large number of indicators, thus increasing the model complexity. Tests for mediation effects can also potentially increase the sample size requirement (Kenny et al., 1998).

Unlike CBSEM, PLS uses an iterative algorithm to separately solve blocks of the measurement model and subsequently estimate the structural path coefficients. This iterative method successively estimates factor loadings and structural paths subset by subset. As such, the estimation procedure employed by PLS allows researchers to estimate highly complex models as long as the sample size is adequate to estimate the most complex block (relationship) in the model. The literature suggests that PLS is appropriate for testing the magnitude of moderation effects (Helm et al., 2010) and for performing between-group comparisons (Qureshi and Compeau, 2009). PLS is more likely to detect between-group differences than CBSEM when data are normally distributed, sample size is small, and exogenous variables are correlated. Thus, we suggest that researchers consider PLS when the research model is extremely complex and may lead to estimation problems in CBSEM.

2.2.3. Data Properties

CBSEM generally requires a multivariate normal distribution of the sample data. Non-normal data may lead to underestimated standard errors and inflated goodness-of-fit statistics in CBSEM (MacCallum et al., 1992), although these effects are lessened with larger sample sizes (Lei and Lomax, 2005). In social science research, data often do not follow a multivariate normal distribution, thus limiting the applicability of CBSEM in some circumstances. Compared with CBSEM, PLS generally places less strict assumptions on data distribution. PLS also does not require a multivariate normal data distribution. Because PLS is regression-based, it generally only requires the data distribution assumptions of the ordinary least squares (OLS) regression. PLS "involves no assumptions about the population or scale of measurement" (Fornell and Bookstein, 1982, p.443) and consequently works with nominal, ordinal, and interval scaled variables.

Therefore, if violations of data distribution assumptions could potentially undermine CBSEM estimation, researchers should consider using PLS. A close examination the results of both CBSEM and PLS provides a useful robustness check of the analysis.

2.2.4. Specifying Formative Constructs

Although the presence of formative constructs does not preclude the use of CBSEM, CBSEM generally lacks the ability to estimate research models with formative constructs. Applying CBSEM to research models with formative constructs often results in unidentified models (Jarvis et al., 2003). This is because using formative indicators in CBSEM implies zero covariance among indicators, and the model can only be solved when it includes a substantial number of additional parameters (MacCallum and Browne, 1993). Because the algorithms

performed in a PLS analysis generally consist of a series of ordinary least squares analyses (Chin, 1998b), identification is not a problem for recursive models (i.e., models without feedback loops). This feature gives PLS an advantage in estimating research models with formative constructs. PLS can estimate research models with both reflective and formative constructs without increasing model complexity (Chin, 1998a; Vinzi et al., 2010). Therefore, Diamantopoulos and Winklhofer (2001) suggest using PLS when formative indicators are present in the research model. Because the presence of formative constructs in the research model typically leads researchers to consider PLS, we include specifying and evaluating formative constructs as a part of our guideline for using PLS.

The fundamental difference between reflective and formative constructs is that the latent variable determines the indicators for reflective constructs whereas the indicators determine the latent variable for formative constructs (see Figure 1). Researchers can refer to Chin (1998b), Diamantopoulos and Winklhofer (2001) and Petter et al. (2007) for in-depth coverage of reflective versus formative constructs.

-- Insert Figure 1 about here --

If the research model includes formative constructs, researchers should carefully consider the conceptual domain of each formative construct and make sure that measurement items capture each aspect and the entire scope of the conceptual domain. Unlike reflective constructs, formative constructs "need a census of indicators, not a sample" (Bollen and Lennox, 1991, p.307). "Failure to consider all facets of the construct will lead to an exclusion of relevant indicators [and] thus exclude part of the construction itself, [therefore], breadth of definition is extremely important to causal indicators [i.e., formative indicators]" (Nunnally and Bernstein,

1994, p.484). Because content validity is particularly important for formative constructs, Petter et al. (2007) suggest making content validity tests a mandatory practice for assessing formative constructs. As part of the effort to establish the content validity of formative constructs, we recommend that researchers conduct a thorough literature review related to the construct's conceptual domain. When literature is not available or does not support the construct validity, qualitative research methods such as expert interviews, panel discussions, and Q-sorting should be used to ensure content validity (Andreev et al., 2009).

Another potential problem is misspecifying a formative construct as a reflective construct. A review of SEM in OM research suggests that 97% of all studies model latent constructs as reflective (Roberts et al., 2010). The authors argue that the small proportion (3%) of studies that model formative constructs under-represents the true theoretical nature of OM constructs. Petter et al. (2007) report that 29% of the studies published in *MIS Quarterly* and *Information Systems Research*, two leading IS journals, have misspecification problems. When a formative construct is specified as a reflective construct, it may lead to either Type I or Type II errors. As a result, the structural model tends to be inflated or deflated (Jarvis et al., 2003). Jarvis et al. (2003) provide a four-point guideline for determining whether a construct should be reflective or formative: (1) direction of causality, (2) interchangeability of the indicators, (3) covariation among the indicators, and (4) nomological network of the indicators.

We use *operational performance* as an illustrative example of a formative construct because it is a multi-dimensional concept that typically includes cost, quality, delivery, and flexibility. In the OM literature, operational performance is modeled as reflective constructs in some studies (e.g., Cao and Zhang, 2011; Inman et al., 2011). However, it is more appropriate to

model operational performance as a formative construct if one follows the guidelines set by Jarvis et al. (2003) and Diamantopoulos and Winklhofer (2001). First, the direction of causality should be from the indicators to the construct because a firm's operational performance is defined collectively by its cost, quality, delivery, and flexibility performance rather than the opposite (Jarvis et al., 2003). Conceptually, researchers cannot expect that an underlying latent construct of operational performance causes cost, quality, delivery, and flexibility performance to all change in the same direction and with the same magnitude. Second, the measurement items of a particular operational performance dimension are not interchangeable with items measuring other performance dimensions. For instance, items measuring manufacturing flexibility cannot be replaced by items measuring cost, quality, or delivery, and vice versa. Third, a change in one performance indicator is not necessarily associated with changes in other indicators. For instance, conceptually, an item measuring flexibility does not have to correlate with an item measuring manufacturing costs. Fourth, with respect to nomological network, one cannot expect that different operational performance items will be impacted by the same set of antecedents or lead to the same set of consequences. Empirical evidence suggests that different antecedents may impact various operational performance dimensions to different extents (Swink et al., 2007). Similarly, the effect of various operational performance dimensions on outcome variables such as business performance can vary considerably (White, 1996).

Because a formative construct by itself is under-identified, researchers should consider including two or more reflective indicators in each formative construct. These reflective indicators are not usually a part of the research model to be tested, but rather are used as an external criterion to assess the formative construct validity (Diamantopoulos and Winklhofer,

2001). The additional reflective indicators and the set of formative items together allow researchers to estimate a multiple indicators and multiple causes (MIMIC) model (Bollen and Davis, 2009; Diamantopoulos and Winklhofer, 2001) to evaluate the external validity of formative constructs. More details about estimating a MIMIC model are provided in subsection 2.3.1.

2.3. Issues to Consider in the Analysis Stage

2.3.1. Measurement Validity Assessment

CBSEM has a set of well-established procedures for evaluating reflective constructs. Researchers can examine item loadings and cross-loadings and assess various measures of construct reliability and validity. Typical measures of construct reliability include Cronbach's Alpha and composite reliability. Convergent validity can be assessed by checking whether the average variance extracted (AVE) of the construct is greater than 0.50 (at the construct level) and the item loadings are greater than 0.70 and statistically significant (at the item level). Discriminant validity is usually examined by comparing the square root of AVE with the correlations between the focal construct and all other constructs. In PLS, researchers can use similar procedures to evaluate the reliability and validity of reflective constructs. Chin (1998b) recommends that researchers examine Cronbach's Alpha, composite reliability, and AVE to assess reflective construct properties. Because OM researchers who have used CBSEM are generally familiar with techniques for assessing measurement models that involve only reflective constructs, our discussion below focuses on techniques for assessing formative constructs.

Although widely accepted standard procedures for evaluating formative construct properties have yet to emerge, researchers generally agree that the criteria used to evaluate

reflective constructs should not apply to formative constructs (Diamantopoulos and Winklhofer, 2001). As Bollen (1989, p.222) notes, "Unfortunately, traditional validity assessments and classical test theory do not cover cause [formative] indicators." Likewise, Hair et al. (2006, p.788) suggest that "because formative indicators do not have to be highly correlated, internal consistency is not a useful validation criterion for formative indicators."

We summarize various procedures for evaluating formative constructs in Table 2. First, researchers should check multicollinearity of formative indicators (items). High multicollinearity suggests that some items may be redundant. To detect multicollinearity, researchers can examine the correlation matrix, the condition index, and the variance inflation factor (VIF). Examining the VIF is a frequently used means of detecting multicollinearity. "General statistics theory suggests that multicollinearity is a concern if the VIF is higher than 10; however, with formative measures, multicollinearity poses more of a problem" (Petter et al., 2007, p.641). Diamantopoulos and Siguaw (2006) suggest a more conservative criterion of VIF at 3.3. Most PLS software packages do not provide VIF outputs. Calculating the VIF of formative items involves an OLS regression with the formative construct score as the dependent variable and all of its formative items as the independent variables. Gefen and Straub (2005) demonstrate how to obtain construct scores and Mathieson et al. (2001) provide a useful example of reporting multicollinearity.

---Insert Table 2 about here---

Petter et al. (2007) suggest that if some of the formative items exhibit high multicollinearity, researchers can (1) model the construct with both formative and reflective items in which highly correlated items are specified as reflective, (2) remove the highly

correlated items, (3) collapse the highly correlated items into a composite index (e.g., Boßow-Thies and Albers, 2010), or (4) convert the construct into a multidimensional construct, in which highly correlated items are specified as reflective indicators of a latent variable that serves as a formative indicator of the original construct. Regarding the second method, researchers should be very careful in deleting formative items and ensure that the conceptual domain of the formative construct will not change if they delete items with high multicollinearity. We suggest that OM researchers theoretically and semantically assess whether the items exhibiting high multicollinearity are redundant, and then follow the guidelines provided by Petter et al. (2007) to deal with multicollinearity among formative items.

Second, researchers should evaluate each formative item's contribution or importance to the formative index (i.e., the formative construct score). A formative index is a composite created by aggregating the formative items of a construct using their respective item weights. This assessment involves examining each formative item's weight, sign, and magnitude (Götz et al., 2010). For formative items, researchers should examine item weight rather than item loading. The item weight should be statistically significant, the sign of the item weight should be consistent with the underlying theory, and the magnitude of the item weight should be no less than 0.10 (Andreev et al., 2009).

Third, researchers should check the external validity of formative constructs. To establish external validity, researchers should typically assess a MIMIC model (Diamantopoulos and Winklhofer, 2001). To conduct MIMIC, researchers should use at least two reflective items that capture the essence of the formative index, as shown in Figure 2a (see example in Diamantopoulos and Winklhofer, 2001). Alternatively, they can create a reflective construct

17

that serves as a "shadow" of the formative construct (i.e., the reflective construct should capture the essence of the formative construct). The MIMIC model can then be estimated using the formative and the shadow construct (Figure 2b, and see example in Cenfetelli and Bassellier, 2009). Note that the MIMIC model should be estimated using CBSEM for each formative construct and its related reflective items or shadow constructs. This is because researchers should examine overall model fit statistics to determine the validity of the formative construct and such statistics are only available in CBSEM. However, the complete research model may still need to be estimated using PLS if the model is under-identified in CBSEM.

---Insert Figure 2 about here---

Nomological validity is manifested in the magnitude and significance of the relationships between the formative construct and other constructs in the research model, which are expected to be strong and significant based on theory and previous research. Several authors suggest testing the nomological validity of a formative construct by correlating its formative items with variables with which the formative construct should theoretically be correlated (e.g., Bagozzi, 1994; Diamantopoulos and Winklhofer, 2001). Ruekert and Churchill (1984) and McKnight et al. (2002) provide examples of nomological validity analysis.

Finally, researchers can examine the discriminant validity of a formative construct. Klein and Rai (2009) propose that for a formative construct, the intra-construct item correlations should be greater than the inter-construct item correlations. Furthermore, formative items should have stronger correlations with their composite construct score than with that of other constructs. We note that these methods for establishing the discriminant validity of formative constructs are

not yet well-established in the literature, and therefore should be adopted at researchers' discretion.

2.3.2. Structural Model Estimation and Assessment

Because PLS does not assume a multivariate normal distribution, traditional parametric-based techniques for significance tests are inappropriate. PLS uses a bootstrapping procedure to estimate standard errors and the significance of parameter estimates (Chin, 1998b). The default setting in the most popular PLS software, *PLS-Graph 3.0*, is to resample 100 times. The default setting for bootstrapping resampling in another popular PLS software, *SmartPLS*, is to resample 200 times. The number of bootstrap samples recommended in the literature has increased. For instance, Chin (1998b) recommends resampling 500 times. Given the computing power available today, as many bootstrapping samples as possible (>500) should be generated. Although increasing the number of bootstrapping samples does not increase the amount of information in the original data, it reduces the effect of random sampling errors that may arise from the bootstrap procedure. Another issue pertaining to bootstrapping is the sample size of each bootstrapped resampling. The sample size of bootstrap resampling is usually set to equal the sample size of the original data from which the bootstrap samples are drawn (Chung and Lee, 2001). Some researchers argue that in certain circumstances the bootstrapping sample size can be smaller than the sample size of the original data, especially when the original sample is large (Andreev et al., 2009, p.8).

Researchers should consider performing bootstrapping using different resampling schemes to verify the results, as in Ahuja et al. (2003) and Rosenzweig (2009). For instance, in Ahuja et al. (2003), the authors used a default bootstrapping resampling setting of 100 times in a

PLS-Graph and verified the results using settings of 250 and 500 times. After performing bootstrapping procedures, several techniques are available for assessing the structural model in PLS.

First, researchers should check the sign, magnitude, and significance of each path coefficient, all of which should be consistent with theory. To evaluate the predictive power of the research model, researchers should examine the explained variance (R^2) of the endogenous constructs. Using R^2 to assess the structural model is consistent with the objective of PLS to maximize variance explained in the endogenous variables. The literature suggests that R^2 values of 0.67, 0.33, and 0.19 are substantial, moderate, and weak, respectively (Chin, 1998b).

Second, researchers can evaluate the effect size of the predictor constructs using Cohen's f^2 (Cohen, 1998). The effect size is computed as the increase in R^2 relative to the proportion of variance that remains unexplained in the endogenous latent variable. According to Cohen (1988), f^2 values of 0.35, 0.15, and 0.02 are considered large, medium, and small, respectively.

Third, researchers can assess the predictive relevance. Chin (1998b) argues "the prediction of observables or potential observables is of much greater relevance than the estimator of what are often artificial construct-parameters" (p.320). Stone-Geisser's Q^2 (Geisser, 1975; Stone, 1974) is often used to assess predictive relevance and can be calculated using the *blindfolding* procedure, which is available in most PLS software packages. If $Q^2 > 0$, then the model is viewed as having predictive relevance.

Fourth, post-hoc power analyses should be conducted to check if the power of the research study is acceptable (>0.80). The effect size, reliability, the number of indicators, or other factors may affect the statistical power of a hypothesis test. Simply applying the "10 times"

rule of thumb may lead researchers to underestimate the sample size requirement in certain situations, such as small effect sizes and low reliability of measurement items. In other words, applying the "10 times" rule of thumb without performing a formal power analysis may lead to hypothesis tests with low power (Goodhue et al., 2006; Marcoulides and Saunders, 2006). Based on the results of a simulation study, Goodhue et al. (2006) argue that the "10 times" rule of thumb for PLS sample size requirement should only be used when effect sizes are large and constructs are highly reliable. Another Monte Carlo simulation by Marcoulides and Saunders (2006) also shows that the sample size requirement to achieve a 0.80 statistical power increases substantially as factor loadings and item inter-correlations decrease. Considering that OM studies tend to have relatively small effect sizes (Verma and Goodale, 1995), a power analysis is particularly needed.

Fifth, although PLS does not provide overall fit statistics, researchers have recently begun to compute Goodness of Fit (GoF) when using PLS (Tenenhaus et al., 2005), which considers the quality of the complete measurement model in terms of average communality (i.e., AVE) and the quality of the complete structural model in terms of average R^2. The average of communality is computed as a weighted average of all of the communalities using weights as the number of manifest variables in each construct with at least two manifest variables.

Finally, we recommend that researchers conduct alternative analyses to check the robustness of the results. Previous empirical research has compared the parameter estimates of the alternative analysis to evaluate whether the results are similar to those generated by the PLS analysis. For instance, Klein and Rai (2009) compare the ordinary least squares (OLS) path analysis results with the PLS results, and Barrosoet et al. (2010) and Vilareset et al. (2010)

compare maximum likelihood CBSEM results with the PLS results. If PLS is chosen mainly because the data distribution assumption is not met, it is helpful for researchers to run CBSEM and compare the results of CBSEM with those of PLS. Even with the violation of data distribution assumptions, the maximum likelihood estimation procedure employed by CBSEM can be quite robust and may still produce reasonably good estimates of the population parameters (Chin, 1995).

2.3.3. Reporting and Interpreting Results

First, researchers should explain in detail their reasons for using PLS. Rather than present the potential advantages of PLS in general, researchers should explain how PLS can help them overcome specific challenges they face that may render CBSEM inappropriate, such as inadequate sample sizes or non-normal data. Researchers should also be careful not to make generalized statements regarding the ability of PLS to estimate research models using small samples that may violate the multivariate normality assumption.

Second, researchers should report the PLS software used. Explicitly reporting the PLS software used enables researchers to replicate previous research, which is important for providing support to worthwhile theories (Tsang and Kwan, 1999).

Third, researchers should adequately report the results needed to assess the predictive power of the research model. Because PLS emphasizes predictive ability, the explained variance (R^2) for all endogenous constructs in the research model should be reported (Hulland, 1999). For the formative constructs, researchers should report item weights, which represent each formative item's contribution to the formative index. "The interpretation of LVs [latent variables] with formative indicators in any PLS analysis should be based on the weights" (Chin, 1998b, p.307).

We also recommend that researchers report not only the statistical significance, but also the confidence interval of structural paths (Streukens et al., 2010). Hypothesis tests using confidence intervals are advantageous because they provide more information about the parameter estimate (Henseler et al., 2009). Shaffer (1995, p.575) notes, "If the hypothesis is not rejected, the power of the procedure can be gauged by the width of the interval." The literature suggests that researchers can use bias-corrected confidence intervals as an appropriate means for testing the significance of the path coefficients estimated by PLS (Gudergan et al., 2008).

Finally, we suggest that researchers report the statistical power of their studies. Although PLS is believed to have the ability to estimate research models with a smaller sample, researchers still should show that the statistical power of the hypothesis tests are adequate, which is typically a concern for studies with small samples.

3. AN ILLUSTRATIVE EXAMPLE OF USING PLS

In this section, we provide an illustrative example of using PLS to estimate a research model that includes both reflective and formative constructs. The research model is presented in Figure 3, in which *operational performance* is modeled as a formative construct, *cross-functional integration* and *trust with suppliers* are the antecedents, and *customer satisfaction* and *market share* are the outcomes. We use data from the third round of the High Performance Manufacturing (HPM) study (Schroeder and Flynn, 2001) to test the research model. The sample size is 266. The measurement items are presented in Tables 3 and 4.

---Insert Figure 3 about here---

---Insert Tables 3 and 4 about here---

We use *SmartPLS 2.0.M3* to estimate our research model. Because the criteria for assessing reflective and formative constructs are different, we assess the two types of constructs separately. The item loadings, composite reliability (CR), and average variance extracted (AVE) of the reflective constructs are shown in Table 3. All item loadings are greater than 0.70 and significant at the 0.001 level, indicating convergent validity at the indicator level. All AVE values are greater than 0.50, suggesting convergent validity at the construct level. All CR values are greater than 0.70, indicting acceptable reliability. The square root of each AVE (shown on the diagonal in Table 5) is greater than the related inter-construct correlations (shown off the diagonal in Table 5) in the construct correlation matrix, indicating adequate discriminant validity for all of the reflective constructs.

---Insert Table 5 about here---

Regarding the formative construct, we examine the formative item weights, multicolinearity between items, discriminant validity, and nomological validity of the formative construct. For each formative item, we examine its weight (rather than its item loading), sign, and magnitude. Each item weight is greater than 0.10 (Andreev et al., 2009) and the sign of the item weight is consistent with the underlying theory (see Table 4). With the exception of "unit cost of manufacturing (Opf1)," all other items are significant at the 0.01 level. In addition, all VIF values are less than 3.3 (Diamantopoulos and Siguaw, 2006), indicating that multicollinearity is not severe. Although Opf1 is not significant at the 0.01 level, this item should be included in the measurement model because conceptually it is an indispensable aspect of operational performance (Petter et al., 2007). To examine the discriminant validity of the formative construct *operational performance*, we compute the average of intra-construct item

correlations for this construct and the average of intra-construct item correlations between this construct and other constructs. We find that the average of intra-construct item correlations is greater than the average of intra-construct item correlations.

We are unable to assess the external validity of the formative construct by performing the MIMIC because the research design of the HPM project does not include additional reflective items or "shadow" reflective constructs that capture the overall operational performance. However, we are able to assess the nomological validity of the operational performance construct by examining the structural paths of its antecedents and outcomes. As Table 6 shows, our results indicate positive and highly significant relationships between operational performance and its two antecedents and two outcomes, indicating the nomological validity of operational performance measures.

The results of the structural model estimate are shown in Tables 6 and 7. We run the structural model using the bootstrap procedure with 200, 500 and 1,000 times of resampling and the magnitude and significance of the structural paths are consistent.

---Insert Tables 6 and 7 about here---

As the t-statistics and 95% confidence intervals indicate, all path coefficients are significant at the 0.01 level. The R^2 of endogenous constructs are 0.129, 0.050 and 0.095 for operational performance, market share, and customer satisfaction, respectively, which do not appear to be very strong (Chin, 1998b). Because the operational performance construct has more than one exogenous construct (i.e., trust and integration), the relative effect sizes (f^2) of the exogenous constructs are calculated using the equation $f^2 = (R^2_{included} - R^2_{excluded})/(1 - R^2_{included})$. The f^2 of trust and integration are 0.034 and 0.064, respectively, which are considered relatively small

effect sizes (Cohen, 1988). Stone-Geisser's Q^2 for endogenous constructs are 0.0416, 0.0316, and 0.0563 for operational performance, market share and satisfaction, respectively, indicating acceptable predictive relevance.

Regarding the overall quality of the research model, we computed the Goodness of Fit (*GoF*) following Tenenhaus et al. (2005). The GOF is calculated as:

$$GOF = \sqrt{\overline{Communality} * \overline{R^2}} = \sqrt{0.0916 * 0.5793} = 0.2303$$

Our sample size of 266 is well above the minimum sample size requirement of 40 as determined by the "10 times" rule of thumb. The most complex block in our model is the formative construct *operational performance*, which has 4 formative indicators. Although the sample size is deemed adequate using the "10 times" rule of thumb, a statistical power analysis is needed to formally determine if the sample size is adequate. We run a power analysis for each structural path and for the largest structural equation (LSE), which is the dependent latent variable (LV) with the largest number of independent LVs influencing it. As Table 6 shows, the power of each path is much greater than 0.80. In our research model, the LSE is the latent construct *operational performance* with two predictors (i.e., trust and integration) in which the smallest effect size (f^2) is 0.034 (see Table 7). For this effect size, our sample size of 266 can achieve a power of 0.768 at the significance level of 0.05 (α), which is only slightly smaller than 0.80.

Finally, we check the robustness of the PLS results. Because our research model includes both reflective and formative constructs, we are unable to run CBSEM and compare PLS results with CBSEM results. Instead, we calculate the average of the items within each construct and

subject these average values to the OLS regression. The OLS regression results are largely consistent with the PLS results (see Table 6).

4. A SUMMARY OF PLS USE IN THE OM LITERATURE

This section reviews PLS use in recent OM literature. This review allows us to identity which aspects of PLS researchers should pay attention to and also serves as the starting point for creating our guideline for evaluating and using PLS. Because PLS is an empirical research method, we consider OM journals that are recognized as publishing relevant and rigorous empirical research. The *Journal of Operations Management* (JOM), *Management Science* (MS), *Decision Sciences Journal* (DSJ), *Production and Operations Management Journal* (POMS), the *International Journal of Operations and Production Management* (IJOPM), the *International Journal of Production Economics* (IJPE), the *International Journal of Production Research* (IJPR), and *IEEE Transactions on Engineering Management* (IEEE) have been cited as those whose missions involve publishing empirical research examining OM topics (Barman et al., 2001; Goh et al., 1997; Malhotra and Grover, 1998; Soteriou et al., 1998; Vokurka, 1996). Our review also covers several major journals in strategy, management, and organization science that sometimes publish research related to operations management, including *Strategic Management Journal* (SMJ), *Academy of Management Journal* (AMJ), and *Organization Science*. Because the use of PLS among business research communities is a relatively recent phenomenon and we want to focus on issues commonly observed in recent OM research, we review articles published from 2001 to 2011. Because MS, DSJ and IEEE Transactions are multi-disciplinary journals with a large OM component, we only review the PLS articles in these three journals that examined OM-related topics.

We perform a key-word search of the titles, key words, abstracts and full texts of the articles in the targeted journals using the following keywords: "partial least squares," "partial-least-squares," "PLS," "formative," "PLS Graph," "PLS-Graph", and "SmartPLS." We focus our search on papers that use PLS as an SEM approach to test empirical research models. Next, each author individually examine the search results to ensure that the articles using PLS are correctly identified and those not using PLS are not included in our review. In total, we found 42 OM-related articles that use the PLS method within the scope of our journal selection and time frame. Our literature review indicates that no articles using the PLS method to examine OM topics were published in POM, AMJ, and Organization Science from 2001 to 2011. Thus, our summary excludes these three journals. The distribution of these articles by journal and year is presented in Table 8. It appears that the number of OM articles using PLS has increased in recent years, particularly since 2007.

We summarize the papers we review in Table 9. Among the 42 articles, 30 explicitly provide a rationale for using PLS. However, the remaining 12 articles do not explain why PLS was chosen. Not unexpectedly, small sample size is the most frequently cited reason for using PLS (n=14), followed by the exploratory or predictive nature of the study (n=11), the use of formative constructs (n=8), non-normal data (n=6), and high model complexity (n=4). Although a small sample size is cited most frequently as the reason for using PLS, only two of the 42 articles perform a power analysis. The median sample size is 126, with a range from 35 to 3,926. Only 13 articles (31%) have a sample size greater than 200.

---Insert Table 9 about here---

The presence of formative constructs is a commonly cited reason for using PLS. Interestingly, although 19 articles use formative constructs, only eight articles state that the use of formative constructs is the reason for using PLS. Among the 19 articles that use formative constructs, three do not perform any analysis on the measurement properties of the formative constructs, and five use techniques for evaluating reflective constructs (e.g., Cronbach's Alpha, composite reliability, and AVE) to assess formative constructs, which is considered inappropriate. Overall, many of the articles we review do not adequately assess the properties of the formative constructs. Seven articles examine formative construct item weights; four evaluate the multicolinearity of the formative measurement items, mostly using the variance inflation factor (VIF); three examine discriminant validity. None of the articles we review evaluates the external validity of the formative construct because no study includes additional reflective items or constructs to capture the formative constructs.

We find that 26 out of the 42 articles report which PLS software is used. *PLS-Graph* is the most popular PLS software, adopted by 19 of the articles. *SmartPLS*, however, is gaining popularity, considering that all six OM articles that use *SmartPLS* were published after 2009. Only one article adopts *Visual PLS*.

Our review identifies 22 articles that report the details of their bootstrapping procedures. We observe that the number of bootstrap samples generated ranges from 100 to 1,500, with the most common number of resampling being 500 (n=11). Two of the articles perform bootstrapping procedures with different rounds of resampling to check the robustness of the results (Ahuja et al., 2003; Rosenzweig, 2009). This is a good practice for checking the robustness of the significance of path coefficients.

With respect to reporting results, each of the articles we review reports the sign, magnitude, and statistical significance of path coefficients. In general, all of the reviewed articles exhibit a good understanding that the objective of PLS is not to estimate overall model fit, but rather to maximize the variance explained of the endogenous variables. Thirty six of the 42 articles report R^2 of the endogenous variables. However, other techniques for evaluating predictive validity are underused. Only six articles report the effect size (f^2) and four report predictive relevance (Q^2). Among the articles we review, Műller and Gaudig (2011) provide a good example of reporting the predictive validity of the research model because they report R^2, f^2 and Q^2.

We note that some of the problems, particularly those related to bootstrapping procedures, evaluating formative constructs and reporting results, could have been avoided if stricter "quality control" mechanisms related to the use of PLS had been enforced during the review process. We recommend that editors and reviewers request contributing authors to follow rigorous standards when using PLS to help improve the rigor of PLS use.

5. DISCUSSION AND CONCLUSION

Our study aims to provide a practical guideline that helps OM researchers evaluate and use PLS. Our study also reviews PLS use in the recent empirical OM literature, which points to the need for a practical guideline for using PLS tailored to the OM audience.

The use of PLS has been growing in the OM literature and will likely gain more popularity. Given the specific challenges empirical OM researchers face, such as the difficulties of obtaining large samples and a lack of well-established scales, PLS can be a potentially useful approach to SEM. Because many OM researchers are unfamiliar with PLS, an OM-specific

guideline that focuses on practical applications rather than the technical details of PLS will be particularly helpful.

The main contribution of our study is to provide a practical guideline for using PLS with detailed illustrative examples from the OM literature. This guideline is expected to help improve the methodology rigor of PLS use in the OM field. A second contribution is that our study presents a review and summary of PLS use in the OM and related fields. Our review helps OM researchers learn from past PLS use and subsequently improve future PLS use.

Although PLS has been used in a variety of research fields, the extent to which it has been used is far less than that of CBSEM in most research fields. Goodhue et al. (2006) assert that it is only in the IS field where PLS has become the dominant approach to SEM. The somewhat limited use of PLS relative to CBSEM in many research fields seems to reflect researchers' general concerns about the weaknesses of the PLS method. Indeed, statistically, CBSEM is superior to PLS in the sense that parameter estimates are unbiased (Chin 1995). Thus, if CBSEM assumptions are met, researchers should strongly consider using CBSEM.

However, we suggest that concerns about PLS should not preclude it as a potential analysis technique because no empirical methodology is perfect. If the assumptions of the PLS method are met and it is used appropriately, it can be a useful data analysis technique. Our position is that OM researchers should consider PLS when CBSEM is unobtainable due to the violations of some key CBSEM assumptions (e.g., sample sizes and sample distribution) or model identification problems. "PLS is not a competing method to CBSEM. Depending upon the researcher's objective and epistemic view of data to theory, properties of the data at hand, or level of theoretical knowledge and measurement development, the PLS approach may be more

appropriate in some circumstances" (Chin, 1998b, p.295). In fact, "CBSEM and PLS are considered as complementary rather than competitive methods, and both have a rigorous rationale of their own" (Barroso et al., 2010, p.432).

Although we argue that OM researchers should not preclude the possibility of using PLS, we oppose accepting PLS as the preferred approach to SEM without a careful assessment of its applicability. OM researchers should be cautious in assessing their model assumptions and data requirements, especially the sample size requirement because it is often cited as the main reason for using PLS. Because "PLS is not a silver bullet to be used with samples of any size" (Marcoulides and Saunders, 2006, p.VIII), researchers should consider a variety of factors and perform power analyses to determine whether the sample size is adequate to support the statistical inference.

As empirical OM researchers start to recognize the potential of PLS, we expect that more OM researchers will seriously consider PLS as a potential SEM method. We hope our study can serve as a useful guideline to help empirical OM researchers evaluate and use PLS.

References

Ahuja, M.K., Galletta, D.F., Carley, K.M., 2003. Individual Centrality and Performance in Virtual R&D Groups: An Empirical Study. Management Science 49(1), 21-38

Anderson, J.C., Rungtusanatham, M., 1994. A theory of quality management underlying the Deming management method. Academy of Management Review 19(3), 472

Anderson, J.C., Rungtusanatham, M., Schroeder, R.O., Devaraj, S., 1995. A path analytic model of a theory of quality management underlying the Deming management method: Preliminary empirical findings. Decision Sciences. Blackwell Publishing Limited 26(5), pp. 637-658

Andreev, P., Hearty, T., Maozz, H., Pliskin, N. 2009. Validating formative partial least squares (PLS) models: Methodological review and empirical llustration *ICIS 2009 Proceedings*.

Bagozzi, R.P. 1994. Structural equation models in marketing research: Basic principles. R.P. Bagozzi, ed. *Basic principles of marketing research*. Blackwell, Oxford, 317-385.

Barman, S., Hanna, M.D., LaForge, R.L., 2001. Perceived relevance and quality of POM journals: A decade later. Journal of Operations Management 19(3), 367-385.

Barroso, C., Carrión, G.C., Roldán, J.L. 2010. Applying maximum likelihood and PLS on different sample sizes: Studies on servqual model and employee behavior model. V.E. Vinci, W.W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares: Concepts, methods and applications*. Springer-Verlag, Berlin, Germany, 427-447.

Bentler, P.M., Chou, C.P., 1987. Practical issues in structural modeling. Sociological Methods and Research 16(1), 78-117

Bollen, K.A., 1989. Structural equations with latent variables. Wiley-Interscience, New York. NY

Bollen, K., Lennox, R., 1991. Conventional wisdom on measurement: A structural equation perspective. Psychological Bulletin 110(2), 305-314.

Bollen, K.A., Davis, W.R., 2009. Causal indicator models: Identification, estimation, and testing. Structural Equation Modeling: A Multidisciplinary Journal 16(3), 498-522.

Boßow-Thies, S., Albers, S. 2010. Application of PLS in marketing: Content strategies on the internet. V.E. Vinci, W.W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares: Concepts, methods and applications*. Springer-Verlag, Berlin, Germany, 589-604.

Cao, M., Zhang, Q., 2011. Supply chain collaboration: Impact on collaborative advantage and firm performance. Journal of Operations Management 29(3), 163-180

Cenfetelli, R.T., Bassellier, G., 2009. Interpretation of formative measurement in information systems research. MIS Quarterly 33(4), 689-707.

Cheung, M.-S., Myers, M.B., Mentzer, J.T., 2010. Does relationship learning lead to relationship value? A cross-national supply chain investigation. Journal of Operations Management 28(6), 472-487.

Cheung, M.-S., Myers, M.B., Mentzer, J.T., 2011. The value of relational learning in global buyer-supplier exchanges: a dyadic perspective and test of the pie-sharing premise. Strategic Management Journal 32(10), 1061-1082

Chin, W.W., 1995. Partial least squares is to lisrel as principal components analysis is to common factor analysis. Technology Studies 2, 315-319.

Chin, W.W., 1998a. Issues and opinion on structural equation modeling. MIS Quarterly 22(1), 7-16.

Chin, W.W. 1998b. The partial least squares approach to structural equation modeling. G.A. Marcoulides, ed. *Modern methods for business research*. Lawrence Brlbaum Associates, Mahwah, NJ, 295-336.

Chin, W.W., Marcolin, B.L., 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(2), 189-217.

Chin, W.W., Newsted, P.R. 1999. Structural equation modeling: Analysis with small samples using partial least squares. R. Hoyle, ed. *Statistical strategies for small sample research*. Sage, Thousand Oaks, CA.

Chung, K.-H., Lee, S.M.S., 2001. Optimal bootstrap sample size in construction of percentile confidence bounds. Scandinavian Journal of Statistics 28(1), 225-239.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*, 2nd ed. Lawrence Erlbaum, Hillside, NJ.

Cording, M., Christmann, P., King, D.R., 2008. Reducing causal ambiguity in acquisition integration: Intermediate goals as mediators of integration decisions and acquisition performance. Academy of Management Journal 51(4), 744-767.

Diamantopoulos, A., Riefler, P., Roth, K.P., 2008. Advancing formative measurement models. Journal of Business Research 61(12), 1203-1218.

Diamantopoulos, A., Siguaw, J.A., 2006. Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. British Journal of Management 17(4), 263-282.

Diamantopoulos, A., Winklhofer, H.M., 2001. Index construction with formative indicators: An alternative to scale development. Journal of Marketing Research 38(2), 269-277.

Fornell, C., Bookstein, F.L., 1982. Two structural equation models: Lisrel and PLS applied to consumer exit-voice theory. Journal of Marketing Research 19(4), 440-452.

Gefen, D., Straub, D., 2005. A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. Communications of AIS 16, 91-109.

Geisser, S., 1975. The predictive sample reuse method with applications. Journal of the American Statistical Association 70(350), 320-328.

Goh, C., Holsapple, C.W., Johnson, L.E., Tanner, J.R., 1997. Evaluating and classifying pom journals. Journal of Operations Management 15(2), 123-138.

Goodhue, D., Lewis, W., Thompson, R. 2006. PLS, small sample size, and statistical power in mis research *the 39th Annual Hawaii International Conference on System Sciences* Kauai, Hawaii.

Götz, O., Liehr-Gobbers, K., Krafft, M. 2010. Evaluation of structural equation models using the partial least squares (PLS) approach. V.E. Vinci, W.W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares: Concepts, methods and applications*. Springer-Verlag, Berlin, Germany, 691-711.

Gudergan, S.P., Ringle, C.M., Wende, S., Will, A., 2008. Confirmatory tetrad analysis in PLS path modeling. Journal of Business Research 61(12), 1238-1249.

Hair, J.F., Black, B., Babin, B., Anderson, R.E., Tatham, R.L. 2006. *Multivariate data analysis*. Prentice Hall, Upper Saddle River, NJ.

Helm, S., Eggert, A., Gardefeld, I. 2010. Modeling the impact of corporate reputation on customer relationship and loyalty using partial least squares. V.E. Vinci, W.W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares: Concepts, methods and applications*. Springer-Verlag, Berlin, Germany, 515-534.

Hennig-Thurau, T., Groth, M., Paul, M., Gremler, D.D., 2006. Are all smiles created equal? How emotional contagion and emotional labor affect service relationships. Journal of Marketing 70(3), 58-73.

Henseler, J., Ringle, C.M., Sinkovics, R.R. 2009. The use of partial least squares path modeling in international marketing. R.R. Sinkovics, P.N. Ghauri, eds. *Advances in international marketing*. Emerald Bingley, 277-320.

Hox, J.J., Maas, C.J.M., 2001. The Accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. Structural Equation Modeling: A Multidisciplinary Journal 8(2), 157-174

Hulland, J., 1999. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. Strategic Management Journal 20(2), 195-204.

Inman, R.A., Sale, R.S., Green Jr, K.W., Whitten, D., 2011. Agile manufacturing: Relation to JIT, operational performance and firm performance. Journal of Operations Management 29(4), 343-355

Jarvis, C.B., Mackenzie, S.B., Podsakoff, P.M., Mick, D.G., Bearden, W.O., 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. Journal of Consumer Research 30(2), 199-218.

Kenny, D. A., Kashy, D. A., Bolger, N. 1998. Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), The handbook of social psychology (Vol. 1, 4th ed., pp. 233-265). Boston, MA: McGraw-Hill.

Klein, R., Rai, A., 2009. Interfirm strategic information flows in logistics supply chain relationships. MIS Quarterly 33(4), 735-762.

Lei, M., Lomax, R.G., 2005. The effect of varying degrees of nonnormality in structural equation modeling. Structural Equation Modeling 12 (1), 1–27.

MacCallum, R.C., Roznowski, M., Necowitz, L.B., 1992. Model modifications in covariance structure analysis: the problem of capitalization on chance. Psychological Bulletin 111 (3), 490–504.

MacCallum, R.C., Browne, M.W., 1993. The use of causal indicators in covariance structure models: Some practical issues. Psychological Bulletin 114(3), 533-541.

Malhotra, M.K., Grover, V., 1998. An assessment of survey research in pom: From constructs to theory. Journal of Operations Management 16(4), 407-425.

Marcoulides, G.A., Saunders, C., 2006. PLS: A silver bullet? MIS Quarterly 30(2), III-IX.

Mathieson, K., Peacock, E., Chin, W., 2001. Extending the technology acceptance model: The influence of perceived user resources. Database 32(3), 86-112.

McKnight, D. H., Choudhury, V., Kacmar, C. 2002. Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research, 13(3), 334-359.

Müller, M., Gaudig, S., 2010. An empirical investigation of antecedents to information exchange in supply chains. International Journal of Production Research 49(6), 1531-1555

Nunnally, J.C., Bernstein, I.H. 1994. *Psychometric theory*, 3rd ed. McGraw-Hill, New York.

Petter, S., Straub, D., Rai, A., 2007. Specifying formative constructs in information systems research. MIS Quarterly 31(4), 623-656.

Qureshi, I., Compeau, D. 2009. Assessing between-group differences in information systems research: A comparison of covariance- and component-based sem. MIS Quarterly 33(1), 197-214.

Roberts, N., Thatcher, J.B., Grover, V., 2010. Advancing operations management theory using exploratory structural equation modeling techniques. International Journal of Production Research 48(15), 4329-4353

Rosenzweig, E.D., 2009. A contingent view of e-collaboration and performance in manufacturing. Journal of Operations Management 27(6), 462-478.

Roth, A.V., Schroeder, R., Huang, X., Kristal, M.M. 2007. Handbook of metrics for operations management: Multi-item measurement scales and objective items.. Sage Publications Newbury Park, CA.

Ruekert, R. W., Churchill Jr, G. A. 1984. Reliability and validity of alternative measures of channel member satisfaction. Journal of Marketing Research, 21(2), 226-233.

Schroeder, R.G., Flynn, B.B., 2001. High Performance Manufacturing: Global Perspectives. Wiley, New York.

Shaffer, J.P., 1995. Multiple hypothesis testing. Annual Review of Psychology 46(1), 561-584.

Shah, R., Goldstein, S.M., 2006. Use of structural equation modeling in operations management research: Looking back and forward. Journal of Operations Management 24(2), 148-169

Soteriou, A.C., Hadijinicola, G.C., Patsia, K., 1998. Assessing production and operations management related journals: The european perspective. Journal of Operations Management 17(2), 225-238.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, Series B 36(2), 111-133.

Streukens, S., Wetzels, M., Daryanto, A., Ruyter, K.D. 2010. Aanalyzing factorial data using PLS: Application in an online complainting context. V.E. Vinci, W.W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares: Concepts, methods and applications*. Springer-Verlag, Berlin, Germany, 567-587.

Swink, M., Narasimhan, R., Wang, C., 2007. Managing beyond the factory walls: Effects of four types of strategic integration on manufacturing plant performance. Journal of Operations Management 25(1), 148-164

Tenenhaus, M., Vinzi, V. E., Chatelin,Y.-M., Lauro, C. 2005. PLS path modeling. Computational Statistics & Data Analysis, 48(1), 159-205.

Tsang, E.W.K., Kwan, K.-M., 1999. Replication and theory development in organizational science: A critical realist perspective. The Academy of Management Review 24(2), 759-780

Verma, R., Goodale, J.C., 1995. Statistical power in operations management research. Journal of Operations Management 13(2), 139-152.

Vilares, M.J., Almeida, M.H., Coelho, P.S. 2010. Comparison of likelihood and PLS estimators for structural equation modeling: A simulation with customer satisfaction data. V. Esposito Vinzi, W.W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields*. Springer, Berlin et al., 289-305.

Vinzi, E., Trinchera, L., Amato, S. 2010. PLS path modeling: From foundations to recent developments and open issues for model assessment and improvement. E. Vinzi, W.W. Chin, J. Henseler, H. Wang, eds. *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields*. Springer, 47-82.

Vokurka, R.J., 1996. The relative importance of journals used in operations management research: A citation analysis. Journal of Operations Management 14(3), 345-355.

Wacker, J.G. 1998. A definition of theory: Research guidelines for different theory-building research methods in operations management. Journal of Operations Management 16(4), 361-385.

White, G.P., 1996. A meta-analysis model of manufacturing capabilities. Journal of Operations Management 14(4), 315-331

White, J.C., Varadarajan, P.R., Dacin, P.A., 2003. Market situation interpretation and response: The role of cognitive style, organizational culture, and information use. Journal of Marketing 67(3), 63-79.

Wold, H. 1966. Estimation of principal components and related models by iterative least squares. P.R. Krishnaiah, ed. *International symposium on Multivariate Analysis*. Academic Press, Dayton, Ohio.

Wold, H. 1982. Soft modeling: The basic design and some extensions. K.G. Jöreskog, H. Wold, eds. *Systems under indirect observations: Causality, structure, prediction, part 2*. North-Holland, Amsterdam, The Netherlands, 1-54.

Table 1: A guideline for evaluating and using PLS

Issues to consider in the pre-analysis stage (2.2)

Should PLS be used as a data analysis method?
1. Research objectives – exploratory study (2.2.1)
2. Sample size and model complexity – Small sample sizes and highly complex research models (2.2.2)
3. Data property – data does not follow a multivariate normal distribution (2.2.3)
4. Does the research model include formative constructs? (2.2.4)

If PLS is used later in the data analysis stage:
5. If formative constructs are involved:
   - Consider using items that summarize the meaning of the formative constructs for subsequent construct validity analysis (2.2.4)
   - Consider using reflective items that capture the essence of the formative construct
6. Consider increasing the number of items per construct for reflective constructs (2.2.5)

Issues to consider in the analysis stage (2.3)

1. Check the validity of formative constructs (2.3.1)
2. Structural model estimation (2.3.2)
   - Properly set up bootstrapping procedures that generate the significance level of parameter estimates
3. Assess the research model (2.3.2)
   - Check the model's explanatory power and predictive validity (R^2, f^2, and Q^2)
   - Perform power analysis and robustness check of the results
4. Report results (2.3.3)
   - Report software used to perform PLS analysis
   - Clearly state the rationales for using PLS (nature of the study, construct formulation and data characteristics)
   - Report item weights of formative indicators and item loading of reflective indicators
   - Report statistical power of the analysis
   - Report statistical significance and confidence interval of structural paths

Table 2: Validity tests of formative constructs

	Aspects of validity	Description	Test	Recommended criterion	Note
Item-level tests	The contribution of each item to the formative construct	Formative item weights should be large and significant	Check the sign, magnitude, significance, range, and average of formative item weights (Klein and Rai, 2009)	When N orthogonal formative items are specified, the ceiling on their average weight is sqrt (1/N) – the average weights should not be too far below the ceiling	The weight, rather than the loading of the formative items should be examined (Chin, 1998b)
	Multicolinearity between items	A high multicolinearity suggests that some indicators may be redundant	Check variance inflation factor (VIF)	A VIF below 3.3 indicates the absence of multicollinearity (Diamantopoulos and Siguaw, 2006)	Researchers should be careful about deleting items because doing so can change the conceptual domain of the construct
Construct-level tests	Nomological validity	The relationship between the formative construct and other theoretically related constructs in the research model should be strong	Check the structural path coefficients related to the formative construct		
	External validity	The formative index should explain the variance of alternative reflective items of the focal construct to a large extent (Diamantopoulos and Winklhofer, 2001)	Check the reflective item factor loadings\n\nEstimate a multiple indicators and multiple causes (MIMIC) model (Bollen and Davis, 2009)	The reflective indicators should have a significant and large factor loading\n\nThe MIMIC model should have a good model fit	Researchers need to should develop reflective items for the formative construct, mainly for checking construct validity.\nMIMIC should be fitted using CBSEM (Diamantopoulos and Winklhofer, 2001)
	Discriminant validity	*Compare item-to-own-construct-correlations with item-to-other-construct-correlations (Klein and Rai, 2009)	Formative items should correlate with their composite construct score to a greater extent than with the composite score of other constructs		

This method was recently proposed in the literature (Klein and Rai, 2009) and is not as well-established as the other validity tests listed in the above table. Klein and Rai (2009) do not provide detailed guidance on how to apply this test.

Table 3: Measurement properties of reflective constructs

Construct	Indicator (label)	Item Loading	T-Stat.	Composite Reliability	Communality (AVE)
Trust with Suppliers	We are comfortable sharing problems with our suppliers (Tst1)	0.8678	28.3379	0.8592	0.6709
	In dealing with our suppliers, we are willing to change assumptions in order to find more effective solutions (Tst2)	0.7851	16.4536		
	We emphasize openness of communications in collaborating with our suppliers (Tst3)	0.7911	16.5213		
Cross-functional Integration	The functions in our plant work well together (Int1)	0.8829	45.7463	0.9180	0.7367
	The functions in our plant cooperate to solve conflicts between them, when they arise (Int2)	0.8450	31.0738		
	Our plant's functions coordinate their activities (Int3)	0.8550	38.1051		
	Our plant's functions work interactively with each other (Int4)	0.8458	29.9594		
Customer Satisfaction	Our customers are pleased with the products and services we provide for them (Sat1)	0.9273	78.6283	0.8998	0.7511
	Our customers seem happy with our responsiveness to their problems (Sat2)	0.7522	14.1533		
	Our customers have been well satisfied with the quality of our products, over the past three years (Sat3)	0.9072	44.6741		
Market Share	How large is the plant's market share, relative to the next largest competitor? For example, a response of 200% indicates your market share is twice that of the next largest competitor (Mrkt)	1.0000	--	--	--

Table 4: Measurement properties of formative constructs

Construct	Indicator	Item Weight	T-Stat.	VIF
Operational Performance	Unit cost of manufacturing (Opf1)	0.1494	1.2695	1.078
	Conformance to product specifications (Opf2)	0.3651	3.7948	1.152
	On time delivery performance (Opf3)	0.5886	5.6621	1.208
	Flexibility to change product mix (Opf4)	0.3269	2.8544	1.118

Table 5: Construct correlations

	X1	X2	X3	X4	X5
Trust with Suppliers (X1)	*0.8191*				
Cross-Functional Integration (X2)	0.2846	*0.8583*			
Operational Performance (X3)	0.2659	0.3077	--		
Market Share (X4)	0.0435	-0.1039	0.2237	--	
Customer Satisfaction (X5)	0.2793	0.2583	0.3088	0.0746	*0.8667*

Note: The square root of average variance extracted (AVE) is shown on the diagonal of the correlation matrix and inter-construct correlations are shown off the diagonal.

Table 6: Structural estimates

Path	PLS Result			OLS Regression Result		Power
	Coefficient	**T-Stat.**	**95% Confidence Interval**	**Coefficient**	**T-Stat.**	
Trust with Suppliers → Operational Performance	0.2043	3.3447	(0.2007, 0.2079)	0.194	2.727	0.9238
Cross-functional Integration → Operational Performance	0.2611	4.2546	(0.2574, 0.2648)	0.252	3.546	0.9926
Operational Performance → Market Share	0.2235	3.6089	(0.2196, 0.2273)	0.224	3.148	0.9613
Operational Performance →Customer Satisfaction	0.3199	5.6398	(0.3165, 0.3233)	0.309	4.451	0.9998

Table 7: R^2, communality, and redundancy

Construct	R^2	Communality (AVE)	Redundancy	Q^2	f^2
Trust with Suppliers	--	0.6709	--	**--**	0.034
Cross-functional Integration	--	0.7367	--	**--**	0.064
Operational Performance	0.1293	0.4115	0.0403	0.0416	--
Market Share	0.0501	N/A	0.0501	0.0316	--
Customer Satisfaction	0.0953	0.7511	0.0678	0.0563	--
Average	**0.0916**	***0.5793**	**0.0527**	**0.0432**	**--**

*The average of communality is computed as a weighted average of all of the communalities using weights as the number of manifest variables in each construct with at least two manifest indicators.
.

Table 8: Distribution of empirical OM articles that use PLS

		DSJ	IEEE	IJOPM	IJPE	IJPR	JOM	MS	SMJ	Total
Year	2000	0	1	0	0	0	0	0	0	1
	2001	0	1	0	0	0	0	0	0	1
	2003	0	0	0	0	0	0	1	0	1
	2004	1	0	0	0	0	1	1	0	3
	2005	0	0	1	0	0	0	0	0	1
	2006	1	0	0	0	0	0	0	0	1
	2007	2	1	0	0	0	2	1	0	6
	2008	0	0	2	0	0	0	1	0	3
	2009	1	0	0	2	0	2	0	0	5
	2010	3	2	2	2	3	1	0	1	14
	2011	0	0	0	2	3	0	0	1	6
Total		8	5	5	6	6	6	4	2	42

The list of PLS articles we reviewed is available upon request.

Table 9: Summary of the OM articles that use PLS (n=42)

Rationales for using PLS 30 articles specify the rationale for using PLS	[a]Breakdown of rationales for using PLS						
	Exploratory or predictive nature of the study	Small sample size	Model complexity	Formative constructs used	Non-normal data	No rationale for using PLS specified	
	*11	14	4	8	6	12	

Sample size	[b]Sample size summary (n)						
	Mean=246	Median=126	Min=35	Max=3,926			
	Sample size distribution						
	n<50	50<n$\leq$100	100<n$\leq$150	150<n$\leq$200	200<n$\leq$300	300<n$\leq$500	n>500
	1	13	11	4	8	4	1

Formative constructs 19 articles use formative constructs	Assessment of formative constructs						
	Contribution of items to the construct	Multicolinearity between items	[c]Nomological validity	External validity	Discriminant validity	Formative constructs not assessed	Formative constructs assessed as reflective constructs
	7	4	N/A	0	3	3	5

Bootstrapping 22 articles report details of bootstrapping procedures	Number of bootstrapping samples						
	n=100	n=200	200<n<500	n=500	n>500		
	1	4	2	11	5		

PLS software used 26 articles report PLS software used	*PLS Graph*	*SmartPLS*	*Visual PLS*				
	19	6	1				

Report results	Statistical power analysis performed	Structural path confidence interval reported	R^2 reported	f^2 reported	Q^2 reported	Formative item weights reported	
	2	0	36	6	4	14	

Each number in the above table indicates the number of articles that are classified in a given category.
[a]*Some articles provide more than one rationale for using PLS.*
[b]*In cases where sample sizes can be counted differently depending on the level of observation, we use the smaller sample size to be conservative.*
[c]*We did not summarize nomological validity because each article reports some statistically significant structural paths, which to some extent demonstrates the nomological validity of formative constructs.*
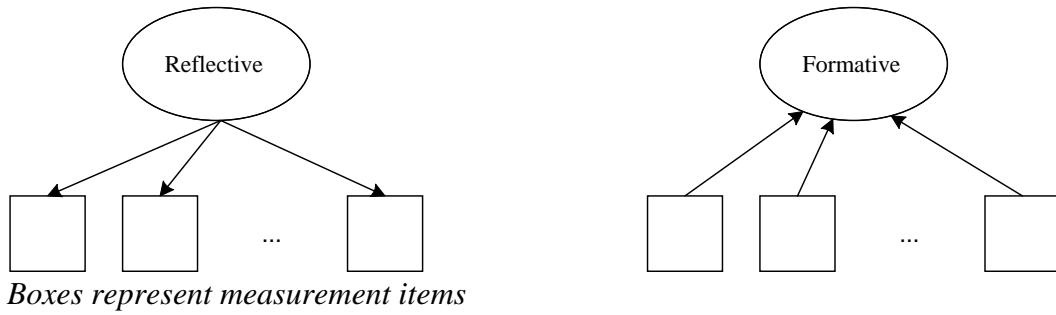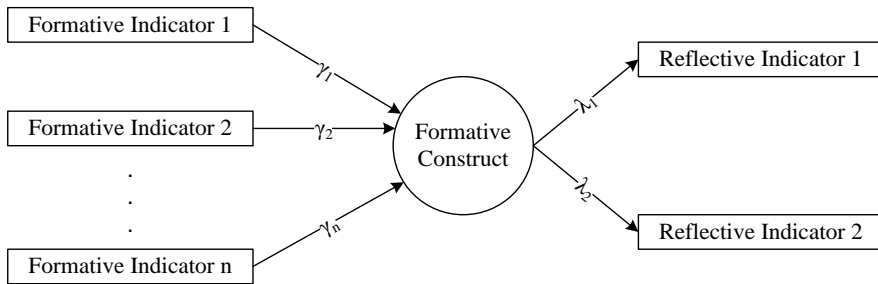
Figure 1: Reflective and formative constructs

Boxes represent measurement items

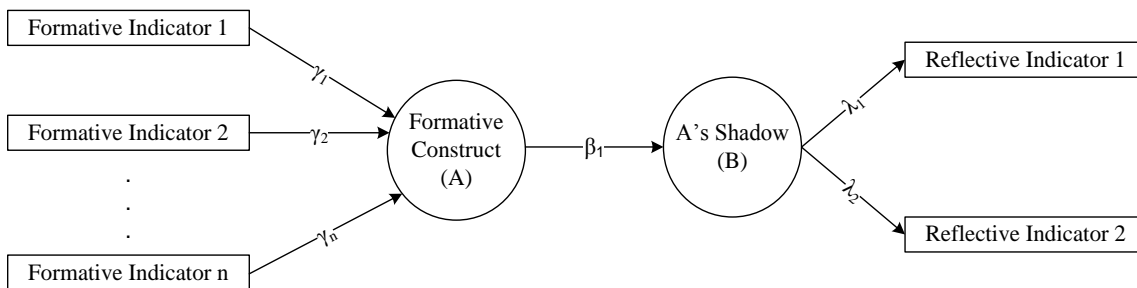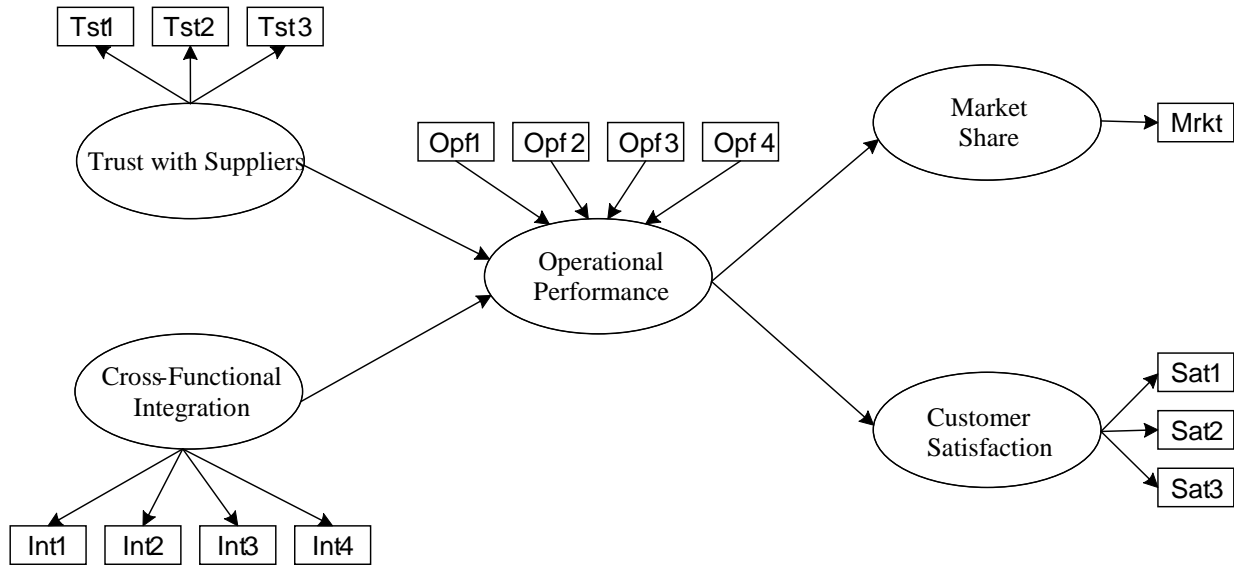# Figure 2: MIMIC tests

Figure 2a

Figure 2b

Figure 3: The illustrative research model

Appendix A: PLS algorithms (*PLS-Graph*)

The basic idea behind the PLS algorithm is relatively straightforward. First, the PLS algorithm uses an iterative process to estimate item weights that link the items to their respective latent variable. Second, once the final item weights are obtained, the latent variable (LV) scores of each LV are calculated as a weighted average of its items. Here the item weights estimated earlier are used as the weights for aggregating item scores into the LV scores. Finally, the LV scores just estimated are used in a set of regression equations to estimate the structural path weights (i.e., relationships between LVs) (Fornell and Bookstein, 1982).

Central to the PLS algorithm is the estimation of item weights, which uses an iterative process that almost always converges to a stable set of item weight estimates. The procedure for obtaining item weights is shown in Figure 4. Each iteration involves a two-step estimation. The two steps are called *inside approximation* and *outside approximation*, respectively. Inside approximation generates LV scores as a weighted average of item scores based on the item weight estimates. Outside approximation generates LV scores as a weighted average of the LV scores of the neighboring LVs based on the structural path weights. In each iteration, the inside approximation first uses LV score estimates from the previous round of outside approximation to calculate structural path weights. The structural path weight between two LVs is equal to the correlation between the two

LV scores if the two LVs are structurally connected, and zero otherwise. Next, PLS uses these structural path weights to compute a new set of LV scores. In the inside approximation, PLS uses the just generated LV scores to estimate a new set of item weights. Finally, these item weights are used to generate another set of LV scores that will be used in the next iteration. The method for determining item weights using factor scores is similar to simple regression for reflective constructs (i.e., the item scores of each reflective item are regressed on the LV scores) and similar to multiple regression for formative constructs (i.e., LV scores are regressed on all of the formative items of the LV). PLS repeats the iteration until the percentage changes of each outside approximation of items weights relative to the previous round are less than 0.001. Once the final item weights are obtained, PLS calculates the LV scores of each LV as the weighted average of its items. PLS then uses the LV scores just generated to estimate the structural path weights using ordinary least squares (OLS) regression. The LV scores of each dependent LV in the research model are regressed on the LV scores of the respective independent LVs.

Figure 4: PLS algorithms
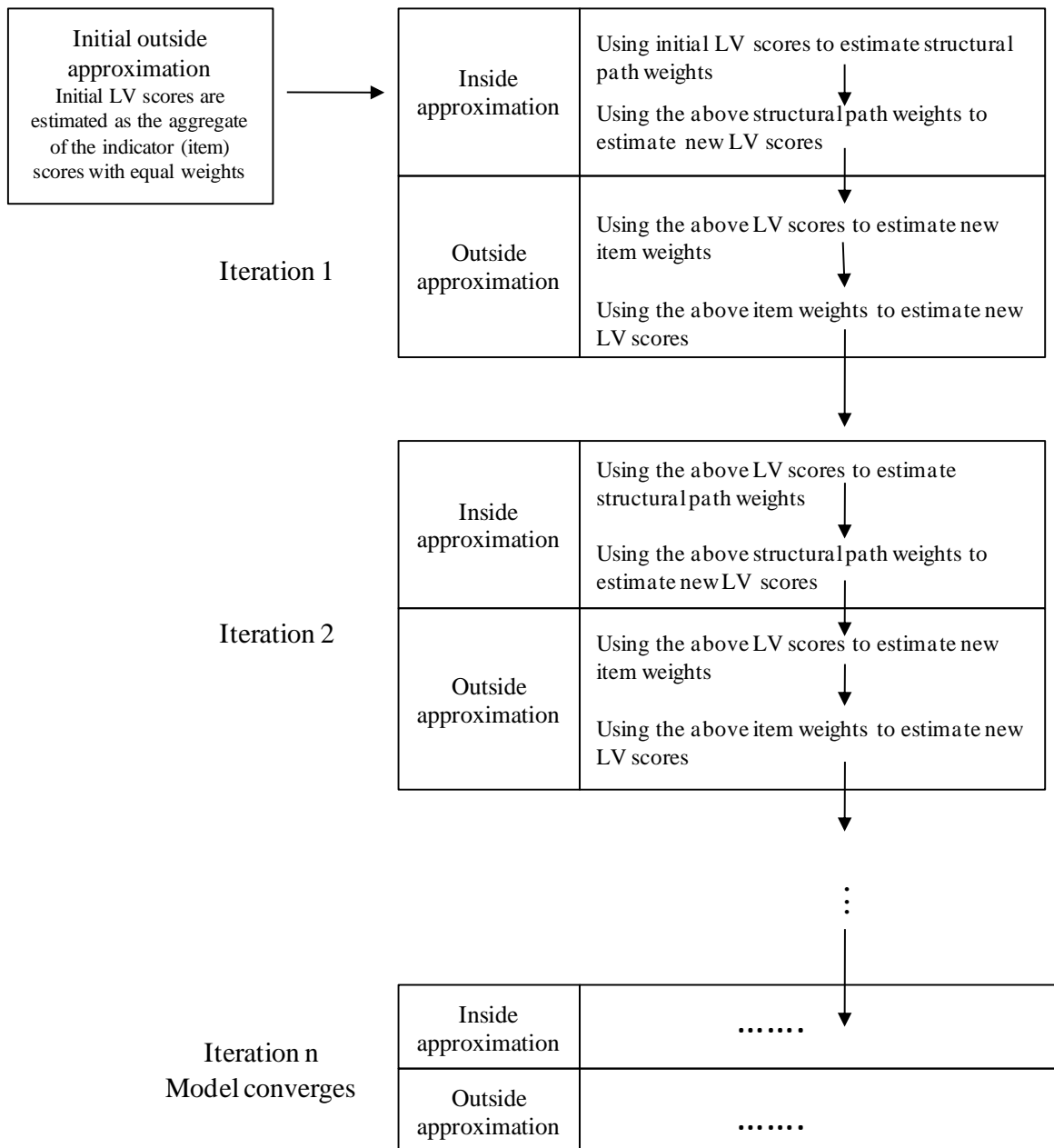
The structural path weight between two connected latent variables is estimated as the correlation between the latent variable (LV) scores of the two LVs.

Item weights are estimated by simple regression for reflective items (item scores of each reflective item are regressed on LV scores) and by multiple regression for formative items (LV scores are regressed on the set of formative items).