

Assessing the predictive performance of structural equation model estimators¹

Joerg Evermann²

Memorial University of Newfoundland

Mary Tate³

Victoria University of Wellington

Revised Version

January 2016

¹ The authors would like to thank the Center for Health Informatics and Analytics at Memorial University of Newfoundland for providing access to the computing facilities that enabled this research.

The complete set of simulation results and scripts used in this research are available from the first author.

² Faculty of Business Administration, Memorial University of Newfoundland, Elizabeth Avenue, St. John's, A1B 3X5, Newfoundland & Labrador, Canada, jevermann@mun.ca, Telephone 1 709 864 8527

³ School of Information Management, Victoria University of Wellington, Box 500, Wellington, New Zealand, mary.tate@vuw.ac.nz

Abstract

Structural equation models are traditionally used for theory testing. With the increasing importance of predictive analytics, and the ability of structural equation models to maintain theoretical plausibility in the context of predictive modeling, identifying how best to predict from structural equation models is important. Recent calls for a refocusing of partial least squares path modeling (PLSPM) for predictive applications further increase the need to assess and compare the predictive power of different estimation methods for structural equation models. This paper presents two simulation studies that evaluate the performance of different modes and variations of PLSPM and covariance analysis on prediction from structural equation models. Study 1 examines all-reflective models using blindfolding and the Q^2 statistic. Study 2 examines mixed formative-reflective models using out-of-sample cross-validation and the RMSE statistic. Recommendations to guide researchers in the choice of appropriate prediction method are offered.

Keywords: Prediction; Structural Equation Models; Partial Least Squares; Simulation Study

Assessing the predictive performance of structural equation model estimators

Introduction

Explanation and prediction are two main purposes of theories and statistical methods (Gregor, 2006). Explanation is concerned with the identification of causal mechanisms underlying a phenomenon. On the statistical level, explanation is primarily concerned with testing the faithful representation of causal mechanisms by the statistical model and the efficient estimation of unbiased parameter values from samples, that is, making valid inferences to population parameters. In contrast, prediction is the ability to predict values for individual cases based on a statistical model whose parameters have been estimated from a suitable training sample.

Quantitative research in management has been dominated by causal-explanatory statistical modeling at the expense of predictive modeling (Shmueli, 2010; Shmueli & Koppius, 2011). The advent of big data has changed this. Modern organizations, not only analytics leaders such as Facebook, Google, Amazon and Walmart, but also smaller and less prominent businesses, are generating petabytes of data that record billions of digital transactions annually (Davenport, 2006, 2013). Carrying out predictive modeling on such large datasets has the potential to generate fresh insights for business practitioners and drive new theorizing for management researchers (Shmueli, 2010; Shmueli & Koppius, 2011).

Structural equation models represent latent and manifest variables and their relationships in a single statistical model. The estimation of such models has traditionally relied on covariance analysis methods, usually with the maximum likelihood (ML) estimator. However, the use of partial least squares path modeling (PLSPM) to estimate such models is increasing in many management disciplines, for example in strategic management (Hair, Sarstedt, Peiper and Ringle,

2012a; Hulland, 1999), marketing (Hair, Sarstedt, Ringle and Mena, 2012b; Henseler, Ringle and Sinkovics, 2009), management information systems (Ringle, Sarstedt and Straub, 2012), operations management (Peng and Lai, 2012) and organizational research (Sosik, Kahai and Piovoso, 2009).

Covariance analysis estimates a structural equation model by minimizing the difference between the model-implied and the observed covariance matrices. Because covariance analysis offers unbiased estimates and provides tests of model fit (Antonakis, Bendahan, Jacquart and Lalive, 2010; Rönkkö and Evermann, 2013), covariance analysis is typically associated with explanatory modeling. In contrast, the PLSPM technique treats the latent variables as weighted composites of their manifest indicator variables and estimates the composite model using multiple regression, resulting in biased parameter estimates. Consequently, PLSPM is often recommended for prediction instead (Hair, Ringle and Sarstedt, 2011; Hair et al., 2012a, 2012b; Henseler et al., 2009; Ringle et al., 2012). Herman Wold, who originally developed PLSPM, positioned PLSPM as a method for prediction (Wold, 1982), de-emphasizing the importance of statistical tests and inference to population parameters. Lohmöller later (1989, pg. 72f) writes about PLSPM that "predictor specification is a shortcut term for the type of model building where the investigator

- Starts with the purpose of prediction
- Sets up a system of relations ...where the structure of the relations must be founded in the substance of the matter, and the predictive purpose should not jeopardize a structural causal interpretation of the relation.
- ...The contrast between predictive vs. structural/causal is not absolute...For simple models both aspects come at the same time; for complex models there is a parting of the ways."

Most recently, prominent PLSPM researchers have called for a re-orientation of PLSPM towards predictive or forecasting applications and its abandonment for explanatory modeling: "We also propose a new 'back-to-basics' research program, moving away from factor analysis models and returning to the original object of constructing indices that extract information from high-dimensional data in a predictive, useful way." (Dijkstra, 2010, pg. 23) "PLS path modeling can and should separate itself from factor-based SEM and renounce entirely all mechanisms, frameworks and jargon associated with factor models. ... A logical candidate for an alternative measurement framework is one that is based on forecasting." (Rigdon, 2012, pg. 348)

The emphasis on prediction is reflected in applied research in the management disciplines. Ringle et al. (2012) report that 15% of PLSPM studies in management information systems and almost a quarter of PLSPM studies in other leading management journals claim to focus on prediction. Hair et al. (2012a) report that more than 30% of PLSPM studies in strategic management appeal to predictive goals. More than one quarter of PLSPM studies in marketing are motivated by predictive goals (Hair et al., 2012b).

The context of structural equation models for prediction raises important questions. In general, a statistical model (not limited to structural equation models) that leads to optimal explanation (minimizing bias) does not necessarily also lead to optimal prediction (minimizing bias and estimation error) (Shmueli, 2010). Consequently, the development of predictive models is primarily driven by data, not theory, to the point that modern prediction methods are entirely atheoretical, eschewing easily interpretable regression models for neural networks, support vector machines, nearest-neighbor methods and others (Hastie, Tibshirani and Friedman, 2009). These considerations naturally raise the question as to the role of theory, and therefore also structural equation models, in prediction, and the general relationship between theory development and

prediction (Shmueli, 2010; Shmueli and Koppius, 2011). For example, is the insistence on a, typically theoretically constrained, structural equation model for prediction, as argued for by Lohmöller (1989) in the above quote, over possibly superior a-theoretical models defensible (Rönkkö, Antonakis, and McIntosh, 2016)? Should researchers take the risk of compromising both prediction and explanation for the pragmatically important interpretability of theoretically plausible models (Davenport, 2013; Freitas, 2013; Huysmans, Dejaeger, Mues, Vanthienen, and Baesens, 2011)? What role does the predictive power of explanatory models play in theory evaluation, selection, and development (Shmueli, 2010; Shmueli and Koppius, 2011)? Moreover, while Shmueli & Koppius (2011) present six ways in which predictive models can contribute to theoretical development, these ways do not imply that the prediction model coincides with the theoretical model, as is the case for prediction from structural equation models considered here.

A thorough discussion of these issues is beyond the scope of this article, which has a narrower focus. Specifically, in light of the arguments about the suitability of PLSPM for predictive purposes (Hair et al., 2011, 2012a, 2012b; Henseler et al., 2009; Rigdon, 2012; Rigdon, 2014; Ringle et al., 2012) combined with the dearth of supporting empirical evidence, *this paper addresses the choice of optimally predictive estimation methods for structural equation models with a focus on the variants of PLSPM and ML estimation*. Two simulation studies evaluate the performance of different PLSPM variants and compare PLSPM based prediction to ML and other methods. Study 1 examines all-reflective models using blindfolding and the Q^2 statistic, as recommended by Chin (2010). Study 2 examines mixed formative-reflective models using cross-validation and the RMSE statistic that are typically used in predictive analytics evaluation (Hastie et al., 2009).

The remainder of the paper is structured as follows. The next section discusses existing work on prediction from structural equation modeling with a focus on PLSPM. The following section presents challenges for prediction from structural equation models, followed by an introduction of the design factors common to both simulation studies. The subsequent two sections present the study design, results, and recommendations for each simulation study. The paper concludes with an overall discussion.

Prior Work

Numerous studies in the past have focused on evaluating and comparing covariance analysis (particularly ML estimation) and PLSPM. However, almost all of them have focused on parameter accuracy (bias) and statistical power. These are key issues in inferential applications, but are not as important for predictive modeling (Shmueli, 2010). In contrast, despite the oft-repeated claims about the advantage of PLSPM for predictive modeling (Hair et al., 2011, 2012a, 2012b; Henseler et al., 2009; Ringle et al., 2012), few studies have systematically tested these claims.

Evermann and Tate (2012) examine the predictive ability of reflective factor models using both PLSPM and ML estimation. Prediction from PLSPM estimated models, judged by the Q^2 statistic on blindfolded data sets, is superior to estimation from ML estimated models. However, their use of reflective exogenous constructs in the factor models precludes out-of-sample evaluation of prediction performance through cross-validation, the accepted standard in the predictive analytics literature (Hastie et al., 2009). Becker, Rai and Rigdon (2013) examine the predictive ability of PLSPM estimated models with formative/composite constructs. While Becker et al. (2013) use cross-validation, they do not focus on the recoverability of individual

scores, but on the R^2 of the regression of the endogenous formative construct. As noted by Sharma, Sarstedt, Shmueli and Kim (2015) and Shmueli, Ray, Velasquez Estrada and Chatla (in press), this statistic is a measure of in-sample explanatory power, not a predictive measure. Moreover, because Becker et al. (2013) use a statistically unidentified model, they cannot compare PLSPM with covariance estimation. Most recently, Evermann and Tate (2014) use a cross-validation approach for mixed formative-reflective models and conclude that PLSPM is superior to ML and linear regression methods in their simulation scenario, where predictive power is assessed as the mean RMSE (root mean square error) across indicators.

An important aspect not examined by previous studies is the performance of different PLSPM estimation methods. Specifically, Evermann and Tate (2012) use only PLSPM mode A estimation, because, as they argue, PLSPM mode A is the accepted way of estimating reflective models in the applied literature. For their later work, Evermann and Tate (2014) use only one combination of mode A and B, again reflecting current practice in the applied literature. Recently, Dijkstra and colleagues (Dijkstra and Schermelleh-Engel, 2013; Dijkstra and Henseler, 2015a, 2015b) have developed a consistent PLS estimator (PLSc) that uses disattenuation by estimated composite reliabilities to correct estimated regression path coefficients, yielding yet another PLSPM variant. Their initial simulation studies focus only on parameter bias and efficiency of estimation, so that the usefulness of PLSc for prediction remains to be explored.

A second aspect that has been neglected is prediction from misspecified models. While one would expect prediction from a model with random misspecifications to be poor, more interesting misspecifications are those that add paths to the model, leading, in the limit, to a fully saturated model. Given the lack of a model fit test for PLSPM (Evermann and Tate, 2010), researchers using PLSPM may be inclined to saturate their model with additional paths. Moreover, because

of the different aims of explanatory and predictive models, underspecified models, trading off bias against variance, may be able to predict better than fully specified models (Hastie et al., 2009; Shmueli et al., in press). Hence, prediction under model misspecification is an important criterion in practice.

Prediction from Structural Equation Models

In the context of predictive modeling, structural equation models can present unique challenges. In predictive modeling, the values for a new case are predicted from the predictor variables for that case. However, many structural equation models in the management disciplines are specified as fully reflective (Ringle et al., 2012; Hair et al., 2012a, 2012b) with all manifest variables as endogenous and only latent variables as exogenous. Hence, no manifest predictors exist from which values for new cases can be predicted (Evermann and Tate, 2012, 2014). This lack of predictors limits the usefulness of these models to guide interventions in business practice. For example, in an application of the Technology Acceptance Model (TAM) (Davis, 1989), a manager may be able to measure the perceived usefulness indicator scores for her employees, but the theoretical model does not provide a prediction path from the perceived usefulness indicators to the behavioral intention indicators. In the terminology of Shmueli et al. (in press), operative prediction is not possible from a fully reflective model.

Similarly, when a model is specified in a purely formative way, as is done by Becker et al. (2013), the model contains no manifest predicted variables; only latent variables are predicted. This type of model makes the assessment of the recoverability of individual scores as an evaluation of predictive performance impossible (Evermann and Tate, 2012, 2014). In the

framework of Shmueli et al. (in press), this model allows only latent prediction, not operative prediction.

One can of course simply disregard the directionality of relationships in the structural equation model, picking any set of variables and deciding that some are predictors and some are predicted. One can then proceed with prediction, but runs the danger that the exercise is irrelevant in the domain of interest: Fully reflective or fully formative models suggest that the substantive theory may not have easily identifiable manifest predictors or predicted variables.

Previous research has dealt with these challenges in different ways. Evermann and Tate (2012) evaluate purely reflective models using a blindfolding technique and make the assumption that manifest variables linked to exogenous latent variables are predictors. For a purely formative model, Becker et al. (2013) choose to not predict individual scores on specific variables, but instead evaluate the R^2 of the regression of an endogenous latent variable, an explanatory but not predictive statistic (Sharma et al., 2015; Shmueli et al., in press). Evermann and Tate (2014) restrict their study to mixed reflective-formative models of the kind shown in Figure 1. These models have both manifest predictors and predicted manifest variables. Specifically, all exogenous latent variables are specified formatively, whereas all endogenous latent variables are specified reflectively. In the example model in Figure 1, x_1 , x_2 , and x_3 clearly are predictors that can be used to predict values for y_1 to y_3 and z_1 to z_3 .

Figure 1 here.

Study Design

Models for simulation studies should be representative of those found in the substantive literature. PLSPM models in highly-ranked management information systems journals have a

median number of 7 to 9 latent variables with 9 to 11 structural relations (Ringle et al., 2012), those in the marketing literature have a median of 7 latent variables with 8 structural relations (Hair et al., 2012b), and those in strategic management have a median of 6 latent variables with 9 structural relations (Hair et al., 2012a). Based on these reports, this study examines the models in Figures 2–4 (indicators are not shown). While models 1 and 2 are relatively simple models, model 3 matches the typical characteristics of PLSPM models in the literature quite well.

Figure 2 here.

Figure 3 here.

Figure 4 here.

PLSPM based studies in management information systems have a median sample size of 198. Those in other business disciplines have a median sample size of 160 (Ringle et al., 2012). Studies in the marketing discipline have a mean sample size of 198 (Hair et al., 2012b) and studies in strategic management have a sample size of 155. Given these reports, and following Evermann and Tate (2012, 2014), sample sizes of 100, 250, and 750 are examined.

Both Ringle et al. (2012) and Hair et al. (2012b) report a median of 3.58 indicators for each reflective construct in management information systems and marketing research; Hair et al. (2012a) report a median of 3 indicators and a mean of 3.4 indicators for each reflective construct in strategic management research. Given these reports, and again following Evermann and Tate (2012, 2014), models with 3, 5, or 7 indicators per construct are examined.

Simulations are conducted under conservative conditions: Manifest variables are continuous from a multivariate normal distribution with no missing values, all structural paths are significant with a regression coefficient of 0.75. Table 1 shows the design factors common to both studies.

Table 1 here.

Study 1: Reflective Models and Blindfolding

The first study examines all-reflective models using the technique of blindfolding to compute the Q^2 statistic (Chin, 2010). The following subsections first present blindfolding and the Q^2 statistic. Next, the experimental design factors are presented, followed by a presentation of results and recommendations for researchers.

Blindfolding and the Q^2 statistic

Because the exogenous variables in a fully reflective model are latent, not manifest variables, out-of-sample evaluation of predictive performance is not possible (Shmueli et al., in press). Instead, PLSPM researchers advocate the use of blindfolding and the Q^2 statistic for assessing the predictive strength of structural equation models (Chin, 2010; Hair et al., 2011, 2012b; Henseler et al., 2009; Ringle et al., 2012; Sarstedt et al., 2014). While blindfolding and use of the Q^2 statistic are the current recommendation for PLSPM research, they are not without problems. Blindfolding is not a true out-of-sample technique (Shmueli et al., in press) and, because the omission sets retain much information about the entire dataset as blindfolding never omits an entire case, may lead to overestimation of predictive ability (Rigdon, 2014; Shmueli et al., press).

In blindfolding, the researcher omits a number of observations from the data set, estimates the model parameters, and uses the estimated model to predict the omitted observations. Blindfolding is applied to a matrix of observations containing N cases and M manifest variables. For an omission distance k , by row and beginning with the first data point (row 1, column 1) of this matrix, every k -th observation is omitted. For an illustration, consider the matrix of $N = 10$ cases and $M = 3$ indicators (x_1, x_2, x_3) shown in the left panel of Figure 5. An omission distance of

$k = 5$ leads to omitting, by row, every 5th observation. This process is repeated k times, yielding k omission sets. The omissions in each set are offset by one datapoint, as shown in the remaining panels in Figure 5. As is evident from the figure, choosing the omission distance as recommended by Wold (1982, p. 33) as a prime number strictly greater than the number of columns and strictly less than the number of rows ($M < k < N$) ensures even omission across rows and columns and ensures that each observation is omitted once. The model is then estimated k times, once with each omission set. To estimate the model, the omitted values in each data set are replaced by the columns means (Chin, 2010).

Figure 5 here.

The estimates for the omitted values are compared to the observed values, using the squared difference:

$$E^2 = \sum_{l=1..k} \sum_{j \in J_l} (\hat{X}_j - X_j)^2$$

Here X_j is the observed value for observation j and \hat{X}_j is the estimated value. The set J_l indexes the omitted values in omission set l . The summation is over all omission sets $l = 1..k$ and over all omitted values in each set J_l .

The differences between the variable mean and the observed values are compared in the same way:

$$O^2 = \sum_{l=1..k} \sum_{j \in J_l} (\bar{X}_j - X_j)^2$$

Here X_j is again the observed value for observation j and \bar{X}_j is the column mean of the column of observation j . The summation is again over all omission sets $l = 1..k$ and over all omitted values in each set J_l .

The predictive measure Q^2 is then calculated as

$$Q^2 = 1 - \frac{E^2}{O^2}$$

One distinguishes communality-based and redundancy-based prediction, with correspondingly differing values for the Q_{comm}^2 and Q_{red}^2 predictive measures. In communality-based prediction, the predicted values are based on the estimated composite scores and the factor loadings. For redundancy-based blindfolding, the composite scores for endogenous latent variables are themselves predicted from the structural model using the estimated regression coefficients. Because the prediction of endogenous composite scores is affected by structural error terms, redundancy-based prediction will necessarily perform worse than communality-based prediction. Therefore, communality-based prediction should be chosen for minimizing the prediction error, but redundancy-based prediction should be chosen when the predictive performance of the structural model is to be evaluated.

Study Design

The three models in Figures 2–4 are estimated using PLSPM mode A estimation, PLSPM mode B estimation, and covariance estimation using ML. For each of these estimation methods, communality-based and redundancy-based prediction from the estimated model is evaluated. While typical PLSPM applications use the centroid inner weighting scheme, this study also examines the path and factor inner weighting schemes. Finally, while the recent development of the consistent PLS estimator (PLSc) (Dijkstra & Schermelleh-Engel, 2013; Dijkstra & Henseler, 2015a, 2015b) is motivated by the goal of to reducing estimation bias of PLSPM, PLSc is included in this study to examine its usefulness also for prediction.

Additionally, the model-based prediction methods are compared to an a-theoretical method to examine whether prediction from a structural equation model provides substantial advantages or disadvantages. Because the absence of manifest exogenous variables rules out linear regression models, the EM algorithm is used. While many a-theoretical methods exist (Hastie et al., 2009), the EM algorithm is familiar to management researchers as a missing value imputation method. The algorithm does not rely on a statistical model, but assumes a multivariate-normal distribution of the observed values. The algorithm estimates the means and covariances of this distribution using the maximum-likelihood method and then samples the missing values from the estimated distribution (Schafer, 1997). Because of this stochastic sampling, following recommended practice, multiple imputation with 20 imputation samples is performed.

Wold (1982, p. 33) recommends that the omission distance k should be a prime number between the number of indicators per composite i and the sample size n . Additionally, Chin (2010, p. 680) recommends a small value of around 5 to 10. Based on these recommendations, the omission distance in this study is the smallest prime that is strictly greater than the number of indicators per composite. The blindfolded block of data, as illustrated in Figure 5, may contain indicators of all or only some latent variables. However, redundancy-based prediction and the Q_{red}^2 metric are applicable only to observations on indicators of endogenous latent variables, whereas communality-based prediction and the Q_{comm}^2 metric can be applied to all manifest variables. Following Chin (2010), only the indicators of endogenous latent variables are blindfolded to be able to compare the relative performance of communality- and redundancy-based prediction for the same data. The Q^2 metric is not reported per indicator or per latent variable because all latent variables have the same number of indicators and all indicators are generated with the same loadings.

Table 2 here.

Table 2 shows further experimental design factors. The models are simulated with strong, medium, and weak measurement loadings. For each experimental condition, 500 samples are estimated (replications). The reported outcome measure is the mean Q^2 over the 500 replications. All computations are performed using the R system (version 3.1.2; R Core Team, 2014). Specifically, the `matrixpls` package (version 0.5.0; Rönkkö, 2014) is used for all PLSPM and PLS_c estimations in this study. Results are doublechecked against the `plpspm` package (Sanchez et al., 2015) and found to be identical to 7 decimal places. The ML estimation is performed using the `lavaan` package (version 0.5-17; Rosseel, 2012). The EM analysis is performed using the `norm` package (version 1.0-9.5). Data are generated by independently drawing the exogenous latent variable true scores from a standard normal distribution and then using structural and measurement equations to compute scores for endogenous variables, adding orthogonal error ϵ as indicated in Table 1.

All ML and PLSPM and PLS_c mode A estimations yielded proper results. EM imputation had significant problems with approximately 20% of solutions in many conditions and did not yield any solutions in many conditions with large proportions of missing data. Problems also occurred with PLS_c mode B estimation where up to 20% of cases yielded clearly inadmissible solutions. Clearly erroneous results are removed and additional cases are simulated to reach 500 cases per experimental condition.

Results – Base Model

Tables 3-5 show the simulation results. The inner weighting scheme does not have any effect on the predictive results (identical to 5 decimal places). Consequently, the Q^2 values reported in

Tables 3-5 are based on the centroid weighting scheme, which is the default in many PLSPM software tools; full results are available from the first author.

Table 3 here.

Table 4 here.

Table 5 here.

Model-based versus EM prediction

The a-theoretical EM-based prediction performs remarkably well across the three different models and experimental conditions. On the other hand, because of the relatively large proportion of missing data compared to a typical missing data situation, the EM algorithm has serious problems in producing a solution, especially for large datasets ($i=7$ and also $i=5$ for model 3).

Mode A versus Mode B Estimation

For PLSPM, the differences between mode A and mode B are not very pronounced, with mean differences in Q^2 of $\Delta=0.0023$ in favor of mode A for redundancy based prediction and $\Delta=0.0014$ for communality based prediction. The differences are more pronounced for PLSc where mode A performs better than mode B for both redundancy based prediction ($\Delta=0.138$) and communality based prediction ($\Delta=0.121$). When considering only the results for medium and large samples, differences between mode A and mode B become negligible for PLSPM prediction (both redundancy and communality), but remain for PLSc (redundancy based $\Delta=0.086$, communality based $\Delta=0.092$), in favor of mode A. Given these findings, the remainder of the analysis focuses on mode A estimation only.

Communality versus Redundancy Prediction

Communality based prediction uses only the composite or factor scores and the loadings for that composite or factor. In contrast, redundancy-based prediction uses the structural model to predict the endogenous composite (factor) scores from the exogenous composite (factor) scores. Its predictions are therefore subject to the errors on the endogenous composites (factors). This difference is evident in the results: As expected, communality-based prediction outperforms redundancy-based prediction for all models, all experimental conditions, and all model-based estimation methods (mean difference $\Delta=0.106$). However, the effect is twice as strong for the PLSc mode A ($\Delta=0.110$) and PLSPM mode A ($\Delta=0.122$) based predictions than for ML based predictions ($\Delta=0.053$).

Redundancy based prediction is the relevant criterion for assessing the predictive ability of the structural model (Chin, 2010; Shmueli et al., in press). Of the redundancy based methods, ML estimation performs better than both PLSPM mode A ($\Delta=0.044$) and PLSc mode A ($\Delta=0.051$). In contrast, when simply the best prediction method is desired, communality based prediction as the superior method should be chosen. However, in that case, the strong performance of the EM algorithm must also be considered.

ML versus PLSc and PLSPM Prediction

PLSPM has been argued to provide superior predictive abilities compared to ML estimation. The results show that this superiority holds for communality based ($\Delta=0.025$) but not for redundancy based prediction, where ML based prediction performs better than PLSPM mode A ($\Delta=0.044$). The lack of superiority for redundancy-based prediction may be a result of the known overestimation of loadings and underestimation of structural paths coefficients by PLSPM.

However, a comparison of PLS mode A, which is intended to correct for this bias, and ML estimation shows that the differences for redundancy based prediction are even more pronounced ($\Delta=0.05$) and that communality based prediction from PLS mode A performs about equally well as from ML ($\Delta=0.006$).

Models and Experimental Conditions

The models in this study have different characteristics. Models 1 and 2 have the same number of latent variables, but whereas three exogenous variables jointly predict two endogenous variables in model 1, a single exogenous variable in model 2 predicts four endogenous variables. Model 3 is more complex, including mediations. ML redundancy based prediction performs better for model 1 than model 2 (mean difference $\Delta=0.108$ over all experimental factors) and model 3 ($\Delta=0.078$). The results for ML communality based predictions are similar but with less pronounced differences ($\Delta=0.070$, $\Delta=0.041$). PLS mode A also predicts better from model 1 than model 2 ($\Delta=0.099$) and model 3 ($\Delta=0.181$) with redundancy based prediction. The differences are again less pronounced for communality based prediction ($\Delta=0.066$, $\Delta=0.001$). The differential behavior of redundancy and communality based prediction with respect to different models is expected as communality based prediction does not take into account the differences in structural models, whereas redundancy based prediction does.

As expected from reliability considerations, predictive ability generally increases as the number of indicators increase. For ML redundancy based prediction, the mean Q^2 statistic over all models and other experimental factors increases by $\Delta=0.081$ when moving from 3 to 5 indicators, and additionally by $\Delta=0.012$ when moving from 5 to 7 indicators. The numbers for ML communality based prediction are similar ($\Delta=0.112$, $\Delta=0.028$). The effect is present but less

pronounced for PLSPM mode A communality based prediction, with an increase in mean Q^2 of $\Delta=0.073$ when moving from 3 to 5 indicators and an additional increase of $\Delta=0.022$ when moving from 5 to 7 indicators. For PLSPM mode A redundancy based prediction, the opposite is true. Here, predictive ability decreases (mean Q^2 difference over all models and other experimental factors $\Delta=-0.025$) when moving from 3 to 5 indicators and further decreases by $\Delta=-0.017$ when moving from 5 to 7 indicators.

Predictive ability behaves as expected with respect to loadings and improves as loadings increase, for all estimation methods and for both prediction methods. The effects are very similar in size, and combined over estimation and prediction methods the mean Q^2 statistic improves by $\Delta=0.067$ when moving from low to medium loadings and improves additionally by $\Delta=0.035$ when moving from medium to large loadings.

Stability

The indeterminacy of factor scores from ML estimation is sometimes cited as a reason for preferring PLSPM (Rigdon, 2012). While different methods for generating factor scores exist, and additionally factors can be rotated, the standard deviations in Table 6 show that once a method is chosen (the `lavaan` package uses the regression method), factor scores are not inherently unstable. The ML standard deviations are of approximately the same size as those for the PLSPM estimations. Dijkstra (2014) points out that the choice of different PLSPM inner and outer modes amounts to the same problem: “The use of a specific proxy cannot take away the inherent and real uncertainty” (pg. 149).

The numerical stability of PLSc mode B based predictions differs from that of the other methods. The standard deviations in Table 6 show that PLSc with mode B outer estimation is

highly unstable in many experimental conditions, in particular with small samples and many indicators ($n=100$, $i=5,7$). Here, standard deviations are one or two orders of magnitude above those for PLSc mode A or PLSPM estimation. This instability occurs despite precautions in the simulation to remove clearly inappropriate solutions. This numerical instability had already been noted by Dijkstra and Schermelleh-Engel (2013) in early studies who suggest that "in small samples c_i [the correction factor] may not be well-defined" (pg. 589). This instability can also be seen in the Q^2 distribution. Figure 6 shows plots of this distribution for model 3 with 5 indicators per latent variable, for different combinations of sample size n and loadings l .

Table 6 here.

Figure 6 here.

Comparison to In-Sample Performance

While blindfolding is not a true out-of-sample evaluation method, blindfolding is also not an in-sample method. Therefore, comparing the predictive performance assessed using blindfolding to the in-sample predictive performance (i.e. no blindfolding) can provide some assessment of the degree of overfitting (Hastie et al., 2009; Shmueli et al., in press). When the in-sample performance is much better than out-of-sample performance, the model is closely fitting the training data, and possibly over-fitting the training data, which can be a problem if the goal is out-of-sample prediction.

Overall, when examining ML, PLSPM mode A, and PLSc mode A estimation, the mean Q^2 difference between blindfolded and in-sample performance across all models, experimental conditions, and prediction methods is $\Delta = 0.0429$, indicating that very little overfitting takes place. Communality-based prediction shows a Q^2 difference of $\Delta = 0.105$ whereas redundancy-

based prediction shows a difference of $\Delta = -0.0196$, indicating the the blindfolded performance is actually better than the in-sample performance. Among the estimation methods, ML shows the largest difference in Q^2 ($\Delta = 0.0825$), followed by PLSc ($\Delta = 0.0267$) and PLSPM ($\Delta = 0.0194$).

Results – Misspecified Models

To examine the performance of the different estimation and prediction methods on overspecified models, structural paths from a to y and from c to x are added to the estimated model 1, and structural paths from a to l and from a to z are added to the estimated model 3. PLSPM mode B is not examined for these models because of its relatively poor performance. Only the results of the ML estimation, PLSc and PLSPM mode A estimation and prediction are reported in Tables 7 and 8. The EM-based predictions do not change as they are model independent and thus are the same as in Tables 3 and 5.

As the structural misspecification leaves the measurement model essentially invariant, the communality based prediction should be the same as for the correctly specified model. Moreover, as the additional paths leave the embedded correct structural model intact and the added paths should be estimated to have zero coefficients, the redundancy-based prediction should be the same as for the correctly specified model. The results in Tables 7 and 8 confirm these expectations. Thus, a researcher cannot gain any advantage by overspecifying or saturating the inner, structural model. On the other hand, such overspecifications also do not negatively affect the predictive performance of the model.

Table 7 here.

Table 8 here.

For an idea as to how much the predictive ability suffers when the estimated model is plainly wrong, the estimated model 3 is changed by removing the paths from k to x and from l to z , and instead adding paths from k to z and from l to x . Table 9 shows the results for communality and redundancy based prediction from ML, PLSc mode A and PLSPM mode A estimated models. Communality-based prediction remains equivalent to that from the correctly specified model, despite the interdependencies between structural and measurement models in the estimation in both the ML and PLS algorithms. As expected, the redundancy-based prediction from ML estimated models suffers considerably for a misspecified structural model. A surprising effect is noticeable for the PLSc and PLSPM mode A based predictions. The redundancy based predictions *improve* over those from the correctly specified model with more indicators ($i=5, 7$ for PLSc; $i=7$ for PLSPM).

Table 9 here.

Because none of the models allow underspecification by removing paths without severely compromising the connectivity of the model, direct paths between a and x , b and y , and c and z are added to the generating model 3, but are omitted from the estimated model, which is therefore underspecified. Because the generating model is different, the Q^2 values are not comparable to previous models, so that Table 10 reports the relative percentage difference between the Q^2 values of the true and of the estimated (underspecified) model.

Table 10 here.

Communality-based prediction suffers little for all estimation methods (ML, PLSc, and PLSPM), with mean differences generally about one half of a percent for ML (-0.443%) and

effectively zero for PLSPM (-0.012%). The number of indicators have an effect on communality-based prediction from PLSPM and PLS_c estimated models. Specifically, the percentage differences in Q^2 are an order of magnitude higher for $i = 3$ (-0.035% for PLPM; -0.420% for PLS_c) than for $i = 5, 7$ (-0.000% for PLSPM; -0.012% for PLS_c).

As expected, redundancy-based prediction suffers significantly, by up to eight percent for ML estimation (mean of -7.60%), and somewhat less for PLS_c and PLSPM estimation (means of -5.05%). The number of indicators also has an effect on redundancy based prediction from PLSPM and PLS_c estimated models, however, opposite to the effect for communality-based prediction. When the number of indicators is small ($i = 3$), redundancy-based prediction suffers more (-6.40% for PLSPM, -7.11% for PLS_c) than for $i = 5, 7$ (-2.39% for PLSPM, -0.91% for PLS_c). No such effect exists for ML estimated models exists.

Recommendations and Discussion

To identify recommendations for researchers from the results for the correct model, the optimal estimation and prediction method is determined for each replication. Classification algorithms (Hall, Eibe, Holmes, Pfahringer, Reutemann and Witten, 2009) are then used to derive model-independent decision rules and decision trees. Cautious about overfitting and in the interest of providing a parsimonious recommendation, the model-independent decision tree in Figure 7, based on the C4.5 algorithm, leads to the optimal method for more than 87% of simulated samples (F-value=0.878).

Figure 7 here.

Similarly, the following decision rule leads to the optimal method for more than 87% of simulated samples (F-value=0.878):


```
If (i <= 3) => EM
If (i <= 5) and (n >= 250) and (l >= 1) => EM
Else => PLSPM Mode A Communality
```

Finally, even the simple rule of choosing PLSPM mode A with communality-based prediction for 5 or more indicators ($i=5, 7$) and EM prediction otherwise leads to the correct choice for more than 83% of simulated samples (F-value=0.833).

One caveat with the decision tree in Figure 7 and the above decision rules is that they appear to select model based prediction predominantly for those cases where EM imputation does not yield a solution, suggesting that, when the data, particularly the proportion of to be predicted data, allow for this, the EM algorithm might also be useful in other conditions. When the EM results are excluded, PLSPM mode A communality based prediction is the preferred method, except for large samples, low loadings, and few indicators, as the following decision rule shows (F-value=0.953):

```
If (n >= 750) and (l <= 0.75) and (i <= 3) => ML Communality
Else => PLSPM Mode A Communality
```

Focusing on the predictive ability of the structural model, and therefore on redundancy based prediction, shows that both PLSc and ML based redundancy prediction are the preferred options, as expressed in the following decision rule (F-value=0.653):

```
If (i <= 3) and (l >= 1.25) and (n <= 250) => PLSc Mode A Redundancy
Else => ML redundancy
```

The results for misspecified models suggest that researchers are not served by simply adding structural paths to and overspecifying their model, possibly until saturation. When paths are added to an already correct model they do not affect the predictive ability at all. On the other hand, when the analysts has reason to doubt the correctness of the structural model, the results

indicate that communality based prediction remains largely unaffected by underspecification and misspecifications, and is the preferred choice.

The way that blindfolding in a reflective model deals with prediction naturally allows for and favors the use of communality based prediction, due to the data set for estimation containing full information for all cases and all manifest variables (recall that the blindfolded dataset replaces omitted values by means) and, barring misspecifications in the measurement model, communality based prediction is not affected by disturbance terms in the structural models, nor is communality-based prediction apparently susceptible to structural misspecification. Despite the formal interdependence of measurement and structural models for ML, PLSc, and PLSPM, the communality-based prediction appears to be, in practice, unaffected by misspecification, indicating effective independence of the measurement models from the structural models. Furthermore, the absence of a test for model fit in PLSPM and PLSc makes no difference in this case as, again barring measurement misspecifications, the correctness of the structural model does not appear to be important to communality based prediction.

In summary, prediction from a fully reflective model is conceptually problematic, as discussed earlier, and the use of blindfolding as a technique to establish predictive ability favors non-model based methods such as EM or, when using theoretically motivated models, favors communality-based prediction. The latter turns out to be effectively not model based as the structural misspecifications appear to have no impact, despite the interdependence between structural and measurement models during model estimation.

Study 2: Mixed Formative-Reflective Models and Cross-Validation

The second simulation study examines models with reflective exogenous and formative endogenous constructs as in the example in Figure 1. In contrast to purely reflective models where no manifest predictors exist, exogenous formative constructs with manifest predictors and endogenous reflective constructs with manifest predicted variables make for a more realistic prediction use case as they allow for operative prediction. The dependent variables for a specific case can be predicted from the known exogenous variables of that case and an estimated model. These types of models also allow the use of linear multiple regression models (LM) techniques, that is, a-theoretical models, for prediction.

Both ML and PLSPM impose considerable constraints on the predictive model by introducing mediating latent variables. Thus, their performance should be worse than that of a linear model that does not impose such constraints: "A path model is ...generally suboptimally predictive ...if the object of the analysis were to predict the response variables, ...we cannot do better than to use a multivariate regression ..." (McDonald, 1996, pg. 266).

A recent simulation study using simple models that do not include latent variables, shows that combining predictors using regression weights is generally sub-optimal to combining predictors using their correlations with the criterion variable (Dana & Dawes, 2004). Dana & Dawes (2004, pg. 328) conclude that "regression coefficients should not be used for predictions unless error is likely to be extremely small by social science standards or sample sizes will be larger than 100 observations to predictors. In other words, regression coefficients should almost never be used for social science predictions." In a simple two-latent model, PLSPM mode A represents correlation weights, whereas PLSPM mode B represents regression weights (Rigdon, 2012). Based on this, Rigdon (2012) argues that PLSPM with mode A should be superior to PLSPM

mode B and LM prediction. However, the models considered in this study contain complex structural models not considered by either Dana & Dawes (2004) in their simulation or Rigdon (2012) in his argument. Moreover, the latent variable model may provide information to the PLS-PM or ML prediction that is not available to LM prediction. Thus, the question as to the relative performance of the different estimation methods remains open.

Cross-Validation

When considering models of the form in Figure 1, the cross-validation method to evaluate predictive ability (Hastie et al., 2009) can be applied. In k -fold cross-validation, a sample of n cases is split randomly into k sub-samples (folds), each with n/k cases. The following procedure is then repeated k times: Select $k - 1$ sub-samples as the training sample and estimate the model using these cases. Using the manifest predictor variables of the remaining testing sub-sample and the estimated model parameters, predict the values of the dependent variables for the testing sub-sample. Note that communality-based prediction is not available as the scores for endogenous latents are not estimated for the testing sub-sample. Prediction in this study is therefore analogous to redundancy-based prediction in study 1 in the sense that the structural model is used for prediction and is evaluated during cross-validation.

Let $\hat{Y}_{i,j,h}$ be the matrix of predicted values for fold i and let $Y_{i,j,h}$ be the matrix of true values for fold i . Each matrix has $j = 1..n/k$ rows and $h = 1..M$ columns, where M is the total number of indicators for all endogenous latent variables. The prediction error is then defined as the root mean squared error (RMSE), averaged over all k folds:

$$RMSE = \frac{1}{k} \sum_{i=1..k} \sqrt{\frac{k}{nM} \sum_{j=1..n/k} \sum_{h=1..M} (\hat{Y}_{i,j,h} - Y_{i,j,h})^2}$$

Study Design

The second simulation uses the same models as the first simulation. All exogenous latent variables have formative indicators (predictors for cross-validation), all endogenous latent variables have reflective indicators (predictands for cross-validation). Sample size n and the number of indicators i are varied as before. Two experimental design factors are added, the correlation between formative indicators of the same construct c and the error (residual) variance for the formative construct e , yielding $3 \times 3 \times 3 \times 3 = 81$ experimental conditions for each of the three models. For each of these conditions, 500 samples (replications) are estimated. Table 11 summarizes the experimental design factors for this study. Data are generated by first constructing a population covariance matrix of the exogenous formative indicators for each latent variable with unit variances and covariance c according to the experimental condition. Formative indicator scores are then sampled from a multi-variate normal distribution with that covariance matrix and summed to form the resulting latent variable true score. Orthogonal error e is added to the latent variable score according to the experimental condition. Data for the endogenous latent variables and indicators are then generated using the structural and measurement model equations.

Table 11 here.

For each sample, the model is estimated using PLSPM with modes A and B and mixed modes AB (mode A for exogenous, mode B for endogenous latent variables) and BA (mode B for exogenous, mode A for endogenous latent variables), ML estimation, and a linear model (LM) that simply regresses the endogenous manifest variables on the exogenous manifest variables. Because the consistent PLS estimator (PLSc) (Dijkstra and Schermelleh-Engel, 2013; Dijkstra and Henseler, 2015a, 2015b) is based on a correction that uses the estimated reliability of

composites under the assumption of a reflective common factor model, this correction, and PLS_c, is not applicable to formative models.

For each sample and prediction method, 10-fold cross-validation is applied, following the recommendation by Hastie et al. (2009, pg. 243). The mean RMSE value over 500 replications is reported as outcome measure. As in study 1, all indicators have the same loadings, and all latent variables have the same number of indicators so that indicator-specific or latent variable-specific RMSE are not reported. All computations are performed using the R system (version 3.1.2; R Core Team, 2014). Specifically, the `pls` package (version 0.4.7; Sanchez et al., 2015) is used for all PLSPM estimations; the ML estimation is performed using the `lavaan` package (version 0.5-20; Rosseel, 2012).

Results – Base Model

Tables 12-14 show the RMSE results for the three models. As in study 1, the inner weighting scheme does not have any effect on the predictive results (identical to 5 decimal places). Consequently, the values reported in Tables 12-14 are based on the centroid weighting scheme which is the default in many PLSPM software tools. The RMSE of the PLSPM mixed mode BA are effectively equal those for PLSPM mode A, and those of PLSPM mixed mode AB are effectively equal those for PLSPM mode B. The mixed mode results are therefore omitted in Tables 12-14; full results are available from the corresponding author. The equality of the mixed modes with modes A or B suggests that only the PLSPM estimation mode for the endogenous latent variables has a significant effect on predictive performance, whereas the PLSPM estimation mode for the exogenous latent variables does not.

To provide an idea of the distribution of RMSE values, Figure 8 shows distribution plots for model 3 for medium sample size ($n=250$) and a medium number of indicators ($i=5$) for varying degrees of structural error variance e and formative indicator covariance c . Table 15 shows the standard deviations of RMSE across all models for the different experimental conditions.

Table 12 here.

Table 13 here.

Table 14 here.

Figure 8 here.

Table 15 here.

The results are generally model independent. The RMSE increases with increasing error variance, which is expected as additional random error in the prediction path will negatively impact the predictive abilities of the model no matter what prediction method is chosen. This effect is weakest for PLSPM estimation and strongest for ML estimation: Coefficients of regressing RMSE on structural error variance e are 0.6821 (PLSPM mode A), 0.6831 (PLSPM mode B), 0.7115 (LM), and 0.8014 (ML). Better prediction is achieved with fewer indicators and larger samples: Regressing RMSE on sample size n and number of indicators i across all models and estimation methods yields regression coefficients of -0.0000386 (sample size) and 0.0028 (number of indicators). This result is not surprising, as a larger sample means more accurate estimations of model parameters and fewer indicators mean that a model with the same number of structural parameters has to predict fewer values.

For PLSPM mode A only, the formative indicator covariance c has a small negative effect on the RMSE (regression coefficient of -0.031), corresponding to a positive effect on predictive performance. The effect is more pronounced as sample size decreases and/or the number of indicators increase.

Overall, the ML-based prediction performs worst, with the largest RMSE values for all models and for all experimental conditions (mean RMSE 0.6179). The PLSPM-based methods generally perform well with small but consistent advantages for PLSPM mode A estimation (mean RMSE 0.5640 for mode A, 0.5711 for mode B). Counter to expectations (McDonald, 1996), all PLSPM methods perform slightly better than even linear multiple regression (mean RMSE 0.5806 for LM).

Comparison to In-Sample Performance

Comparing the out-of-sample predictive performance to in-sample performance is useful to assess the degree of overfitting (Hastie et al., 2009; Shmueli et al., in press). When the in-sample performance is much better than out-of-sample performance, the model is closely fitting the training data, and possibly over-fitting the training data, which can be a problem if the goal is out-of-sample prediction.

Across all experimental conditions and all models, the in-sample RMSE is lower only by $\Delta = 0.0238$. Differences in RMSE are smallest for PLSPM mode A ($\Delta = 0.00784$) and mode B ($\Delta = 0.0123$), followed by ML estimation ($\Delta = 0.0357$) and LM ($\Delta = 0.0391$). A regression of experimental factors on RMSE differences (main effects only) shows that differences are more pronounced for smaller samples and lower formative indicator covariance. Full in-sample results are available from the first author.

The small size of these differences, relative to the standard deviations in Table 15 and the differences between experimental conditions, show that the cross-validated out-of-sample performance is very close to in-sample performance, indicating little evidence of overfitting and thus high generalizability of predictions to new samples.

Results – Misspecified Models

As with the first simulation study, examining the predictive ability of misspecified models is informative. The same misspecifications as in the first simulation study are examined in this study. Structural paths from a to y and from c to x are added to model 1, and structural paths from a to l and from a to z are added to model 3. The additional paths leave the embedded correct structural model intact. Because PLSPM mode A performs better than PLSPM mode B and mixed mode for the correctly specified models, only ML and PLSPM mode A are examined.

Tables 16 and 17 show the mean RMSE values for the misspecified models 1 and 3. Comparing these to the RMSE values for the correctly specified models in Tables 12 and 14 shows that, for both models and medium or large samples ($n=250, 750$), ML based prediction and PLSPM based prediction from the misspecified model are equivalent to predictions from the correctly specified model. For small samples ($n=100$), both ML based and PLSPM mode A based prediction from the misspecified model is slightly worse than that from the correctly specified model. However, the relative performance does not change: PLSPM mode A estimation continues to offer better prediction than ML estimation in all experimental conditions.

Table 16 here.

Table 17 here.

Similar to the first simulation study, to give an idea as to how much the predictive ability suffers when the estimated model is plainly wrong, model 3 is changed by removing the path from k to x and from l to z , and instead adding paths from k to z and from l to x . Table 18 shows the results for prediction from ML and PLSPM mode A estimated models. As expected, the RMSE rises significantly over the RMSE for prediction from the correct model, in some experimental conditions by 50%, for both methods.

Table 18 here.

Again, similar to the first study, direct paths between a and x , b and y , and c and z are added to the generating model 3, but not the estimated model 3. The latter is therefore underspecified. As for the first study, because the generating model is different, the RMSE values are not comparable to the previous models, so that Table 19 reports the relative percentage difference between the RMSE values of the true and of the estimated (underspecified) model.

Table 19 here.

The RMSE is higher for the underspecified model than the correct one in all experimental conditions and for both estimation methods. The mean percentage differences are 14.1% for PLSPM mode A and 12.1% for ML estimation, indicating poorer prediction from underspecified models. The RMSE increases more strongly for smaller structural errors. While the increase is independent of sample size for PLSPM mode A, the RMSE differences for ML estimation show a sample size dependency. At all sample sizes, the RMSE differences for ML estimation are smaller than those of PLSPM mode A estimation for small structural errors ($e = 0$) and approach those for PLSPM-A as structural error increases ($e = 0.4$). However, as the sample size

increases, the RMSE differences for ML estimation also increase for small structural errors ($e = 0, e = 0.1$). Despite an effect of indicator covariance c on the RMSE for correctly specified models (Table 14), no such effect is present for the RMSE differences in the underspecification case.

In summary, irrespective of any experimental factor or estimation method, the underspecified model considered here suffers least in predictive performance when structural error is already high, that is, when prediction performance is already poor.

Recommendations and Discussion

Similar to study 1, recommendations for researchers are identified using classification algorithms (Hall et al., 2009) to derive model-independent decision rules and decision trees. Cautious about overfitting and in the interest of providing a parsimonious recommendation, the model-independent decision tree in Figure 9, based on the C4.5 algorithm, leads to the optimal method for more than 83% of simulated samples (F-value=0.819).

Figure 9 here.

Even the following, much simpler rule leads to the optimal method for more than 81% of simulated samples (F-value=0.816):

If ($c = 0$) => PLSPM mode B

Else => PLSPM mode A

However, considering that the differences in RMSE between PLSPM mode A and PLSPM mode B for $c = 0$ are minimal (Tables 12-14) and that in practice, an indicator covariance of exactly zero is unrealistic even for carefully constructed indicators and carefully collected data, researchers should always use PLSPM mode A for estimation and prediction.

The predictive advantage of PLSPM over LM holds only for correctly specified models, or models that contain the correctly specified models and additional structural paths. Unfortunately, model misspecification is a condition that cannot be identified in PLSPM. Therefore, researchers should use ML and PLSPM estimation in a complementary way, using a two-step approach. As ML estimation offers a statistical test of model correctness, researchers should use ML estimation in a first step to ensure model correctness so that the subsequent prediction from that model with PLSPM is optimally predictive as well as theoretically sound. In the second step, researchers should then estimate and predict from the model using PLSPM mode A. In this application, covariance SEM will estimate population parameters, whereas the PLSPM estimation will result in sample parameters for prediction from a particular sample, with no claim of inference to the population. In contrast to causal modeling, when covariance estimation is used in this context, only model fit is important. Parameter significance and the validity and reliability of the measurement model are of secondary importance, only in support of any secondary explanatory research aims. For the subsequent PLSPM estimation, predictive validity assessment is important, but, again, parameter significance or measurement model validity and reliability considerations are not.

The good performance of PLSPM compared to LM is contrary to the assertion by McDonald (1996) that linear multiple regression is the optimal prediction method. As noted earlier, two rival explanations exist. First, as Rigdon (2012) suggests, the use of correlation weights in the PLSPM mode A algorithm may offer better performance than regression weights (LM). On the other hand, PLSPM mode B also performs better than LM. Second, the structural model may provide additional information that is not available to the LM estimation. On the other hand, such

information is also available to ML estimation, which does not perform as well. Further research is needed to disentangle and identify the individual effects.

The effect of indicator covariance on PLSPM mode A estimation is also unexpected. Increased collinearity (covariances) leads to increased standard errors of the estimated regression parameters as different combinations of regression weights of formative indicators produce the same composite. However, since any combination of weights is equivalent in determining the composite scores, this uncertainty should not affect the predictions derived from the estimated model, as observed for LM, ML, and PLSPM mode B estimation. While small sample sizes and a large number of indicators appear to be the condition in which PLSPM is most susceptible to capitalization on chance (Rönkkö & Evermann, 2013), capitalization on chance is not a satisfactory explanation, as capitalization on chance occurs primarily with a weak structural model (Rönkkö & Evermann, 2013), which is not the case here, and also occurs for PLSPM mode B estimation, whereas indicator covariance has no effect on predictive ability for PLSPM mode B.

Overall Discussion

The main contribution of this study is an evaluation of the predictive ability of competing methods for estimating structural equation models: PLSPM, PLSc, ML, and a-theoretical methods. This study contributes to the ongoing discussion of the relative merits of PLSPM and covariance based methods by showing that PLSPM has a place in the methodological toolbox of management researchers for predictive modeling. Specifically, PLSPM and its new variant PLSc outperform covariance based methods for prediction from structural equation models across many of the conditions examined in this simulation study, and should be the preferred choices

when the research aim is prediction. In contrast to a-theoretical prediction methods, PLSPM and PLS_c allow the researcher to work with an explanatory, theory-based model, to aid in theory development, evaluation, and selection (Shmueli & Koppius, 2011).

The present work should only be seen as an initial study, limited in several ways. First, this study does not evaluate prediction from correlations, as noted earlier. Second, the fact that all latent variables have equal numbers of indicators and all indicators have equal loadings significantly simplifies the experimental design but represents a limitation to be addressed in future studies. Third, as PLSPM and related methods are undergoing active development, further research is required to examine the predictive performance of such novel methods as NSC-PM (non-symmetrical component-based path modeling) developed by Dolce (2015), Dolce, Esposito Vinzi & Lauro (2015) and RGCCA (regularized general component correlation analysis) developed by Tenenhaus & Tenenhaus (2011). Fourth, further work is required to examine the range of options for generating factor scores in covariance-based SEM, for example extensions to Bartlett's method (Yuan and Hayashi, 2010; Yung and Yuan, 2013), in order to determine the optimally predictive method.

The indeterminacy of factor scores from ML estimation is sometimes cited as a reason for preferring PLSPM (Rigdon, 2012). While different methods for generating factor scores exist, and additionally, factors can be rotated, Dijkstra (2014) points out that the choice of different PLSPM inner and outer modes amounts to the same problem: "The use of a specific proxy cannot take away the inherent and real uncertainty" (pg. 149). The standard deviations in Table 6 show that, once a method for generating factor scores is chosen, (this study uses regression scores), factor scores are as stable as the composite scores produced by PLSPM.

Echoing Hair et al. (2012a, 2012b), Ringle et al. (2012) and Sarstedt et al. (2014), the final recommendation to researchers who motivate their choice of PLSPM by appealing to the predictive aim of their study, is to act accordingly. Specifically, researchers should perform blindfolding or cross-validation and report Q^2 or RMSE values (Sarstedt et al., 2014, Shmueli et al., in press). Reporting the endogenous R^2 , as is common current practice, is insufficient (Sharma et al., 2015; Shmueli et al., in press). Additionally, the theoretical motivation, parameter significance testing, and the assessment of validity and reliability of the measures should take a limited and sub-ordinate role to the presentation of appropriate blindfolding or cross-validation procedures and metrics. Researchers need to make an explicit case for how the prediction of specific cases is an appropriate and relevant goal for their study, as opposed to the identification or testing of causal mechanisms, and at the same time need to outline their reasons for using a theoretically-based structural equation model for prediction, rather than an entirely a-theoretic, and possibly more predictive, method. Given the current execution and reporting of many PLSPM based studies, with an emphasis on theoretical motivation, measurement validity and reliability testing, and parameter significance tests, the claims of predictive goals are difficult to accept for many published PLSPM studies.

References

- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly*, *21*, 1086—1120.
- Becker, J.-M., Rai, A., & Rigdon, E. (2013). Predictive validity and formative measurement in structural equation modeling: Embracing practical relevance. In *Proceedings of the 32nd International Conference on Information Systems (ICIS), Milan, Italy*.

- Chin, W. (2010). How to write up and report PLS analyses. In E. Esposito Vinzi, W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares* (pp. 655—690). Berlin: Springer-Verlag.
- Dana, J. & Dawes, R.M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, 29, 317-331.
- Davenport, T. (2006). Competing on analytics. *Harvard Business Review*, 84, 98—107.
- Davenport, T. (2013). Keep up with your quants. *Harvard Business Review*, 91, 120—123.
- Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319—340.
- Dijkstra, T. (2010). Latent variables and indices: Herman Wold's basic design and partial least squares. In E. Esposito Vinzi, W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares* (pp. 23—46). Berlin: Springer-Verlag.
- Dijkstra, T. (2014). PLS' Janus face – Response to professor Rigdon's 'rethinking partial least squares modeling: In praise of simple methods', *Long Range Planning* 47, 146-153.
- Dijkstra, T., & Henseler, J. (2015a). Consistent and asymptotically normal PLS estimator for linear structural equation models. *Computational Statistics & Data Analysis*, 81, 10—23.
- Dijkstra, T., & Henseler, J. (2015b). Consistent partial least squares path modeling. *MIS Quarterly*, 39, 297-316.
- Dijkstra, T., & Schermelleh-Engel, K. (2013). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, 79, 585—604.
- Dolce, P. (2015). *Component-based Path Modeling Open Issues and Methodological Contributions*. PhD Thesis. Università degli Studi di Napoli Federico II, Italy.

- Dolce, P., Esposito Vinzi, V., & Lauro, C. (2015). Non-symmetrical component-based path modeling. In Abdi, H., Esposito Vinzi, V., Russolillo, G., Saporta, G., & Trinchera, L., (editors), *Springer Proceedings in Mathematics & Statistics from PLS'14 - 8th International Conference on Partial Least Squares and Related Methods*.
- Evermann, J., & Tate, M. (2010). Testing models or fitting models? Identifying model misspecifications in PLS. In *Proceedings of the 31st International Conference on Information Systems (ICIS), St. Louis, MS*.
- Evermann, J., & Tate, M. (2012). Comparing the predictive ability of PLS and covariance analysis. In *Proceedings of the 33rd International Conference on Information Systems (ICIS), Orlando, FL*.
- Evermann, J., & Tate, M. (2014). Comparing out-of-sample predictive ability of PLS, covariance, and regression models. In *Proceedings of the 35th International Conference on Information Systems (ICIS), Auckland, New Zealand*.
- Freitas, A. (2013). Comprehensible classification models - A position paper. *ACM SIGKDD Explorations Newsletter*, 15, 1—10.
- Gregor, S. (2006). The nature of theory in Information Systems. *MIS Quarterly*, 30, 611—642.
- Hair, J., Ringle, C., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19, 139—151.
- Hair, J., Sarstedt, M., Pieper, T., & Ringle, C. (2012a). The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. *Long Range Planning*, 45, 320—340.

- Hair, J., Sarstedt, M., Ringle, C., & Mena, J. (2012b). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science*, *40*, 414—433.
- Hall, M., Eibe, F., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin, Germany: Springer Verlag.
- Henseler, J., Ringle, C., & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *20*, 277—319.
- Hulland, J. (1999). Use of partial least square (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal*, *20*, 195—204.
- Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., & Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems*, *51*, 141—154.
- Lohmöller, J. (1989). *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg, Germany: Physica Verlag.
- McDonald, R. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, *31*, 239—270.
- Peng, D., & Lai, F. (2012). Using partial least squares in operations management research: A practical guideline and summary of past research. *Journal of Operations Management*, *30*, 467—480.
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Rigdon, E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, 45, 341—358.
- Rigdon, E. (2014). Rethinking partial least squares: Breaking chains and forging ahead. *Long Range Planning* 47, 161-167.
- Ringle, C., Sarstedt, M., & Straub, D. (2012). A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, 36, iii–xiv.
- Rönkkö, M. (2014). *matrixpls: Matrix-based Partial Least Squares Estimation*. URL: <https://github.com/mronkko/matrixpls> R package version 0.5.0.
- Rönkkö, M., Antonakis, J. & McIntosh, C. (2016). Partial least squares path modeling – Time for some serious second thoughts. Unpublished manuscript.
- Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, 16, 425—448.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1-36.
- Sanchez, G., Trinchera, L., & Russolillo, G. (2015). *plspm: Tools for Partial Least Squares Path Modeling (PLS-PM)*. R package <http://CRAN.R-project.org/package=plspm>
- Sarstedt, M., Ringle, C.M., Henseler, J., & Hair, J.F. (2014). On the emancipation of PLS-SEM: A commentary on Rigdon (2012). *Long Range Planning* 47, 154-160.
- Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Boca Raton, FL: CRC Press/Chapman & Hall.
- Sharma, P. N., Sarstedt, M., Shmueli, G., & Kim, K. H. (2015). Predictive model selection in partial least squares path modeling. In *2nd International Symposium on Parital Least Squares Path Modeling, Seville (Spain)*.

- Shmueli, G. (2010). To explain or to predict? *Statistical Science* 25, 289-310.
- Shmueli, G., & Koppius, O. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35, 553—572.
- Shmueli, G., Ray, S., Velasquez Estrada, J.M. & Chatla, S.B. (in press). The elephant in the room: Evaluating the predictive performance of partial least squares (PLS) path models. *Journal of Business Research*.
- Sosik, J., Kahai, S., & Piovosio, M. (2009). Silver bullet or voodoo statistics? A primer for using the partial least squares data analytic technique in group and organization research. *Group & Organization Management*, 34, 5—36.
- Tenenhaus, A. & Tenenhaus, M. (2011) Regularized generalized canonical correlation analysis. *Psychometrika*, 76, 257-284.
- Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. Jöreskog, S. Wold, & H. Wold (Eds.), *Systems under indirect observation: Causality, Structure, Prediction*. North Holland: Elsevier Publishing Company.
- Yuan, K.-H., & Hayashi, K. (2010). Fitting data to model: Structural equation modeling diagnosis using two scatter plots. *Psychological Methods*, 15, 335 – 351.
- Yung, Y.-F., & Yuan, K.-H. (2013). Bartlett factor scores: General formulas and applications to structural equation models. In R. E. Millsap, L. A. van der Ark, D. M. Bolt, & C. M. Woods (Eds.), *New Developments in Quantitative Psychology* (pp. 385–401). Springer New York volume 66 of *Springer Proceedings in Mathematics & Statistics*.

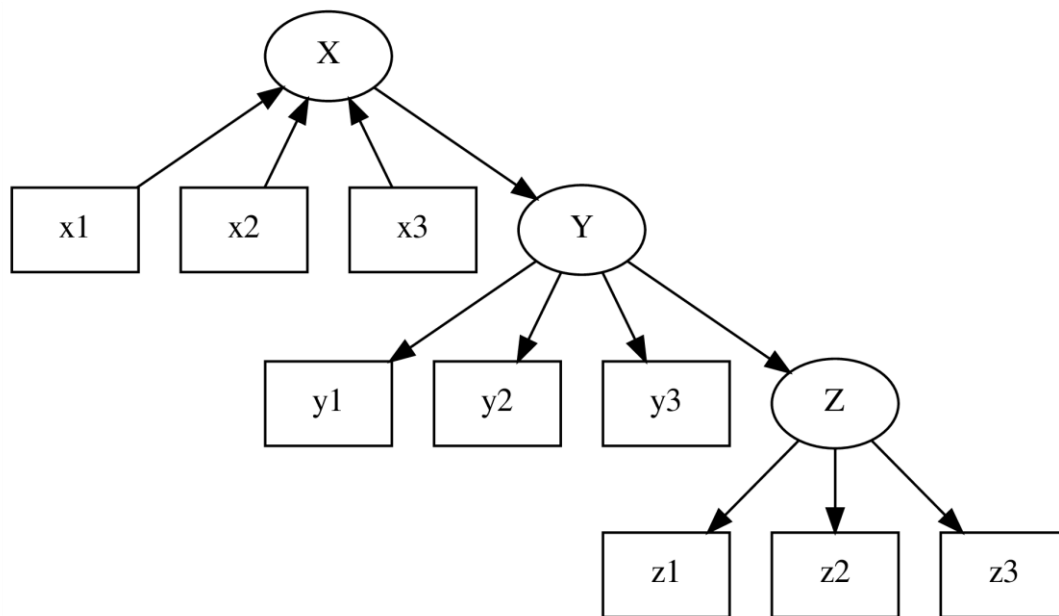


Figure 1: Exogenous formative construct with manifest predictor variables

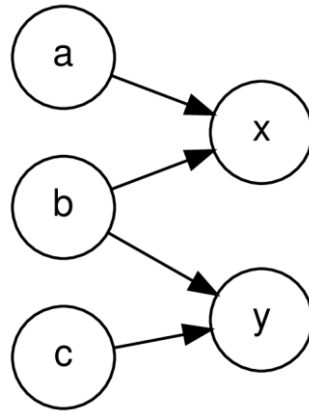


Figure 2: Model 1

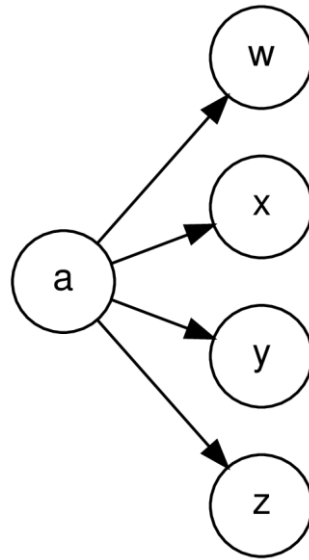


Figure 3: Model 2

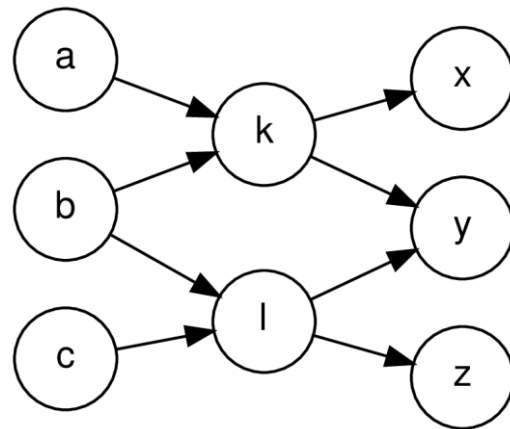


Figure 4: Model 3

X1	X2	X3	X1	X2	X3	X1	X2	X3	X1	X2	X3	X1	X2	X3
	5	4	3		4	3	5		3	5	4	3	5	4
4	3		4	3	3	4	3	3		3	3	4		3
3	4	3		4	3	3		3	3	4		3	4	3
2		4	2	7		2	7	4	2	7	4		7	4
4	6	2	4	6	2		6	2	4		2	4	6	
	5	5	3		5	3	5		3	5	5	3	5	5
1	4		1	4	3	1	4	3		4	3	1		3
3	4	2		4	2	3		2	3	4		3	4	2
2		4	2	6		2	6	4	2	6	4		6	4
5	7	3	5	7	3		7	3	5		3	5	7	

Omission set for $l = 1$ Omission set for $l = 2$ Omission set for $l = 3$ Omission set for $l = 4$ Omission set for $l = 5$

Figure 5: Illustration of omission during blindfolding with $N = 10$ cases, $M = 3$ manifest variables, omission distance $k = 5$ showing the five iterations $l = 1..5$ of omitting values.

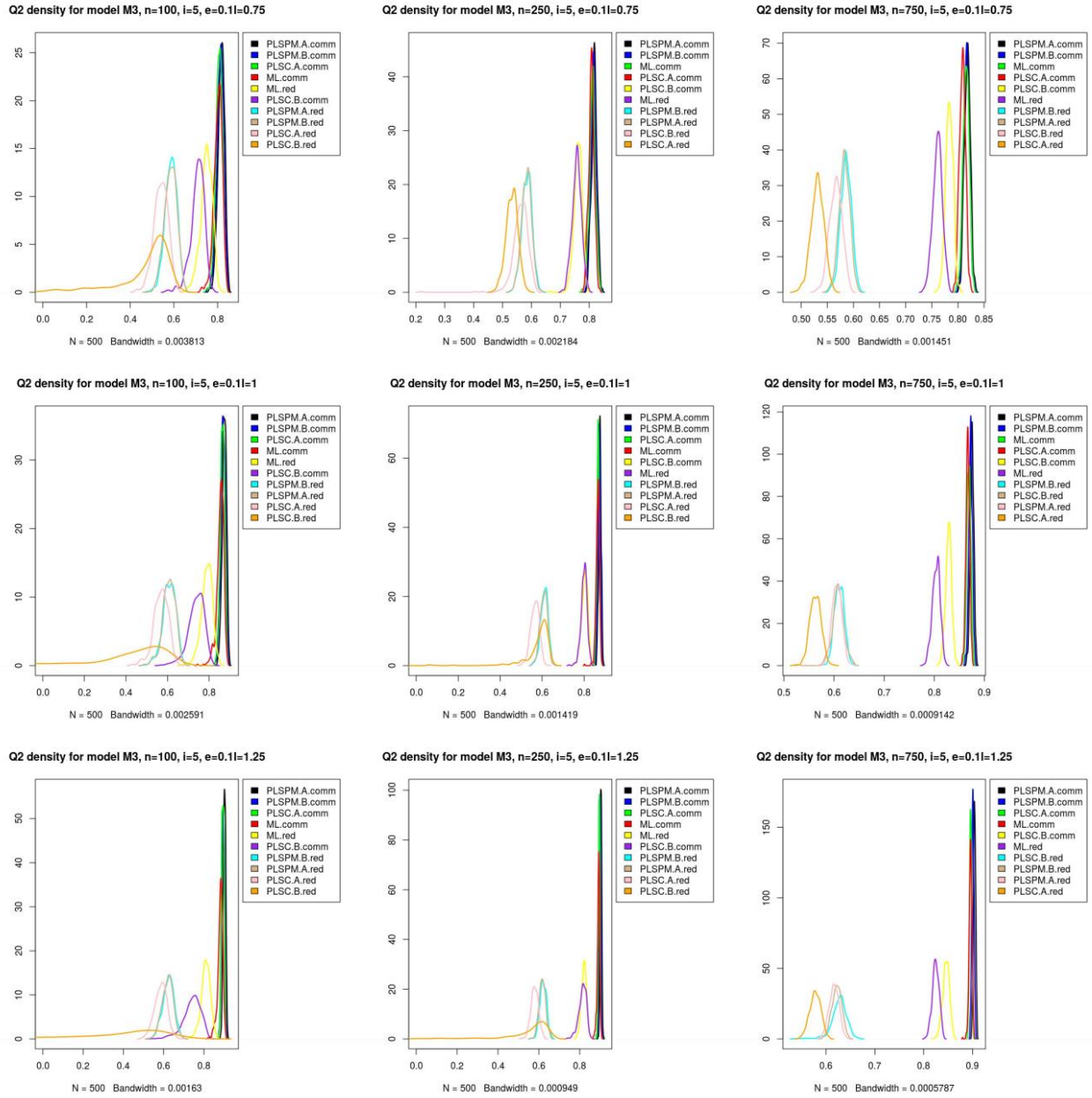


Figure 6: Q^2 distribution plots for model 3, $i = 5$ varying in loadings l and sample size n .

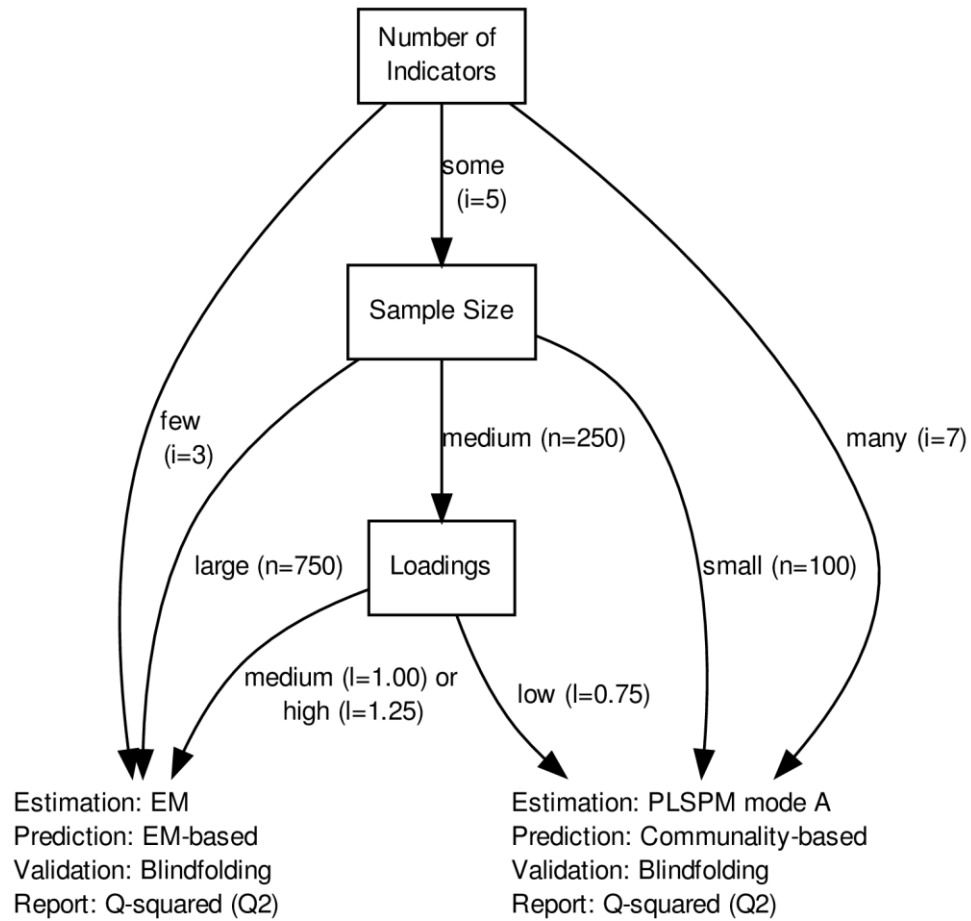


Figure 7: Decision tree for choosing estimation and prediction method for reflective models

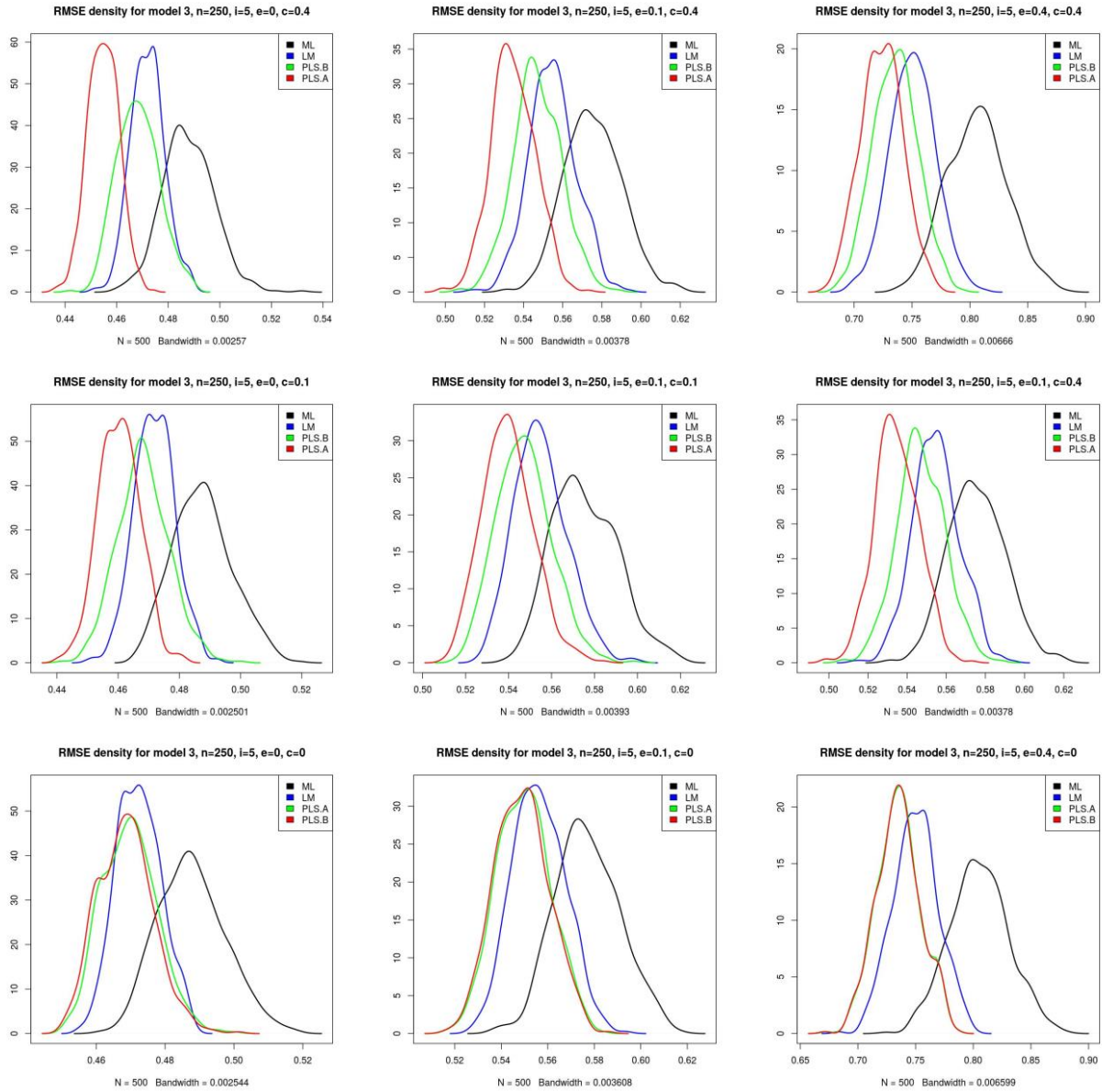


Figure 8: RMSE distribution plots for model 3, $n = 250$, $i = 5$ varying in structural error variance e and formative indicator covariance c .

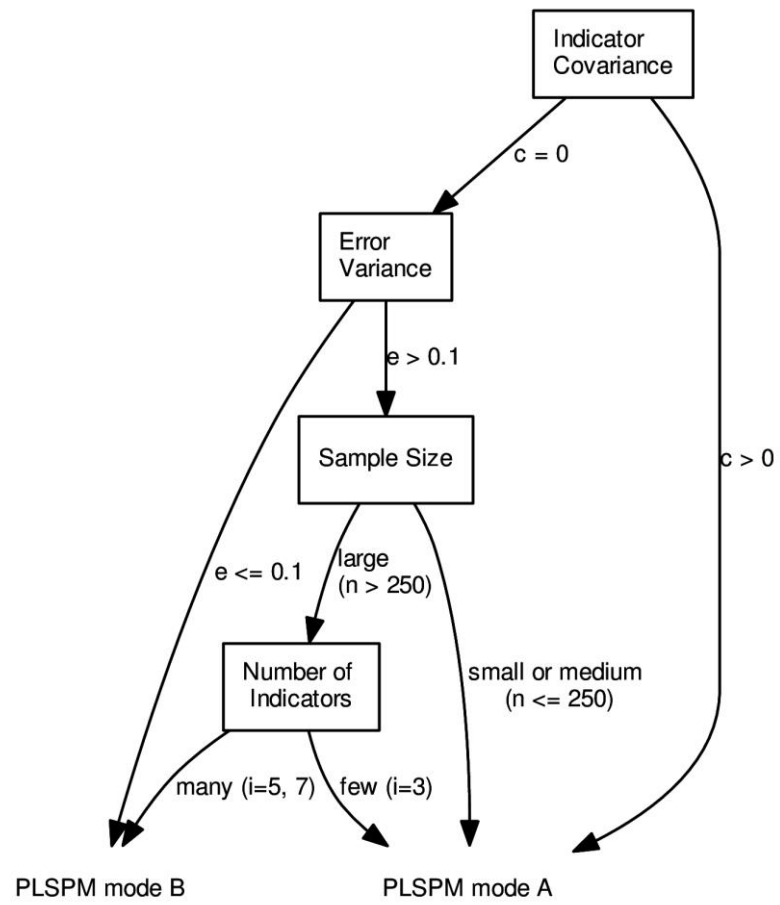


Figure 9: Decision tree for choosing estimation method for formative models

Table 1: Common Experimental Design Factors

Sample size	n	100, 250, 750
Number of indicators per latent construct	i	3, 5, 7
Structural regression coefficients	β	0.75
Error variance on reflective indicators and endogenous latent variables	ε	0.1
Sampling distribution		Normal
Response type		Continuous
Missing values		None

Table 2: Experimental Design Factors for Study 1

Factor loadings (unstandardized)	<i>l</i>	Low (0.75), medium (1), high (1.25)
Estimation methods		PLSPM, PLSc, ML, EM
Prediction methods (for PLSPM, PLSc, ML)		Communality-based, Redundancy-based
Outer modes (for PLSPM)		Mode A, Mode B
Inner schemes (for PLSPM)		Centroid, Path, Factor

Table 3: Mean Q^2 for model 1. Some EM imputations did not converge, indicated by “—”.

All values significantly different from next lowest in row (paired sample t-test, $p=0.05$), except

where noted by *

n	i	l	EM	PLSc						PLSPM				
				ML		Mode A		Mode B		Mode A		Mode B		
				comm	red	comm	Red	comm	Red	comm	red	comm	red	
100	3	0.75	0.779	0.736	0.734*	0.717	0.734	0.692	0.596	0.748	0.730	0.745	0.718*	
		1	0.862	0.787	0.784	0.774	0.792	0.746	0.607	0.803	0.790	0.801	0.775*	
		1.25	0.907	0.816	0.812	0.805	0.822	0.776	0.551	0.833	0.821	0.831	0.807	
	5	0.75	0.772*	0.797	0.769	0.807	0.757	0.717	0.514	0.817	0.754	0.812	0.733	
		1	0.860	0.850	0.815	0.864	0.810	0.764	0.422	0.873	0.809	0.868	0.787	
		1.25	0.909	0.878	0.839	0.894	0.837	0.788	0.404	0.903	0.836	0.897	0.813	
	7	0.75	—	0.826	0.786	0.835	0.770	0.606	0.321	0.839	0.768	0.830	0.738	
		1	—	0.880	0.832	0.891	0.822	0.613	0.236	0.895	0.821	0.885	0.787	
		1.25	—	0.908	0.854	0.920	0.847	0.611	0.231	0.923	0.846	0.913	0.813	
	250	3	0.75	0.810	0.748*	0.745	0.719	0.738	0.710	0.695	0.748	0.732	0.748	0.727
			1	0.882	0.800	0.796	0.777	0.795	0.766	0.730	0.804	0.792	0.804	0.787
			1.25	0.921	0.828	0.823*	0.807	0.825	0.796	0.731	0.833	0.823	0.833	0.818
5		0.75	0.821	0.806	0.778*	0.808	0.758	0.778	0.702	0.817	0.755	0.816	0.747	
		1	0.891	0.860	0.826	0.866	0.812	0.832	0.732	0.874	0.810	0.873	0.801	
		1.25	0.928	0.888	0.849	0.895	0.839	0.860	0.663	0.903	0.838	0.902	0.829	
7		0.75	—	0.833	0.793	0.835	0.771	0.757*	0.670	0.839	0.769	0.837	0.757	
		1	—	0.888	0.839	0.892	0.823	0.798	0.676	0.896	0.821	0.893	0.809	
		1.25	—	0.916	0.862	0.921	0.849	0.815	0.592	0.924	0.848	0.921	0.834	
750		3	0.75	0.820	0.752	0.749*	0.720	0.739	0.717	0.724	0.748	0.733	0.748	0.731
			1	0.889	0.806	0.801	0.778	0.797	0.774	0.776	0.805	0.793	0.805	0.791
			1.25	0.925	0.833	0.828	0.807	0.827	0.804	0.799	0.834	0.824	0.834	0.822
	5	0.75	0.836	0.810	0.782	0.808	0.759	0.799	0.738	0.818	0.756	0.817	0.753	
		1	0.901	0.865	0.830	0.867	0.813	0.857	0.787	0.875	0.811	0.875	0.808	
		1.25	0.934	0.893	0.854	0.896	0.841	0.886	0.810	0.904	0.839	0.904	0.836	
	7	0.75	—	0.837	0.797	0.836	0.772	0.813	0.737	0.840	0.769	0.839	0.766	
		1	—	0.893	0.844	0.893	0.824	0.865	0.782	0.897	0.823	0.896	0.819	
		1.25	—	0.920	0.867	0.922	0.850	0.891	0.803	0.925	0.849	0.924	0.845	

Table 4: Mean Q^2 for model 2. Some EM imputations did not converge, indicated by “—”.

All values significantly different from next lowest in row (paired sample t-test, $p=0.05$), except

where noted by *

n	i	l	EM	PLSc						PLSPM				
				ML		Mode A		Mode B		Mode A		Mode B		
				comm	red	comm	red	comm	red	comm	red	comm	red	
100	3	0.75	0.632*	0.637*	0.617*	0.621	0.624	0.597	0.595	0.652	0.627	0.649	0.627	
		1	0.761	0.715	0.686	0.711	0.706	0.684	0.674	0.742	0.707	0.738	0.708	
		1.25	0.837	0.758	0.725	0.761	0.752	0.731	0.716	0.791	0.753	0.787	0.754	
	5	0.75	0.600	0.703	0.642	0.710	0.638*	0.636	0.572	0.722	0.639	0.717	0.641	
		1	0.748	0.787	0.708	0.800	0.712	0.704	0.630	0.811	0.712	0.805	0.715	
		1.25	0.830	0.834	0.745*	0.850	0.754	0.743	0.663	0.860	0.754	0.854	0.757	
	7	0.75	—	0.730	0.652	0.738	0.640	0.572	0.500	0.744	0.640	0.736	0.643	
		1	—	0.817	0.719	0.828	0.710	0.605	0.526	0.833	0.710	0.823	0.713	
		1.25	—	0.864	0.755	0.876	0.749	0.606	0.526	0.880	0.748	0.870	0.752	
	250	3	0.75	0.687	0.650	0.628*	0.622	0.625	0.614	0.615*	0.652	0.628*	0.651	0.627
			1	0.799	0.728	0.700	0.714	0.709	0.703	0.697	0.744	0.711*	0.742	0.711
			1.25	0.860	0.769	0.735	0.762	0.751*	0.751	0.739	0.791	0.752	0.789	0.753
5		0.75	0.702	0.712	0.651	0.712	0.638	0.682	0.612	0.722	0.640	0.720	0.641	
		1	0.811	0.797	0.719	0.802	0.712	0.765	0.683	0.812*	0.713	0.810	0.714	
		1.25	0.872	0.843	0.754*	0.851	0.752	0.809	0.719	0.860	0.752	0.858	0.754	
7		0.75	—	0.740	0.662	0.741	0.640	0.672	0.584	0.746	0.642	0.743	0.643	
		1	—	0.825	0.728	0.830	0.710	0.735	0.635	0.834	0.711	0.830	0.712	
		1.25	—	0.872	0.763	0.878	0.748	0.759	0.655	0.881	0.748	0.877	0.750	
750		3	0.75	0.707	0.657	0.635	0.625	0.627	0.622	0.624	0.654	0.630	0.654	0.630
			1	0.810	0.733	0.704	0.715	0.708	0.711	0.705	0.743	0.710	0.743	0.710
			1.25	0.868	0.774	0.740	0.763	0.751	0.759	0.748	0.791	0.753	0.790	0.753
	5	0.75	0.729	0.718	0.657	0.713	0.640	0.704	0.631	0.724	0.642	0.723	0.642	
		1	0.828	0.801	0.723	0.803	0.712	0.791	0.703	0.813	0.713	0.812	0.714	
		1.25	0.883	0.848	0.759	0.852	0.752	0.838	0.742	0.861	0.753	0.860	0.753	
	7	0.75	—	0.743	0.665	0.741	0.639	0.718	0.620	0.746	0.641	0.745	0.642	
		1	—	0.829	0.733	0.831	0.710	0.798	0.686	0.834	0.711	0.833	0.712	
		1.25	—	0.877	0.769	0.879	0.748	0.839	0.719	0.882	0.749	0.881	0.750	

Table 5: Mean Q^2 for model 3. Some EM imputations did not converge, indicated by "—".

All values significantly different from next lowest in row (paired sample t-test, $p=0.05$), except

where noted by *

n	i	l	EM	PLSc						PLSPM				
				ML		Mode A		Mode B		Mode A		Mode B		
				comm	red	comm	red	comm	red	comm	red	comm	red	
100	3	0.75	0.727	0.600	0.562	0.667	0.671	0.230	0.212	0.738	0.673	0.741	0.672	
		1	0.830	0.646	0.596	0.718*	0.717	0.168	0.135	0.794	0.708	0.802	0.712	
		1.25	0.884	0.671	0.613	0.744	0.740	0.130	0.096	0.822	0.724	0.835	0.731	
	5	0.75	—	0.802	0.749	0.807	0.541	0.708	0.409	0.817	0.587	0.812	0.585	
		1	—	0.854	0.790	0.865	0.568	0.738	0.327	0.874	0.607	0.868	0.608	
		1.25	—	0.882	0.809	0.893	0.586	0.741	0.245	0.902	0.620	0.895	0.623	
	7	0.75	—	0.826	0.758	0.834	0.514	0.656	0.269	0.839	0.543	0.833	0.537	
		1	—	0.880	0.799	0.890	0.538	0.649	0.179	0.894	0.559	0.887	0.556	
		1.25	—	0.909	0.821	0.920	0.558	0.629	0.158	0.923	0.575	0.915	0.574	
	250	3	0.75	0.798	0.609	0.570	0.667	0.673	0.238	0.255	0.738	0.672	0.746	0.677
			1	0.876	0.657	0.606	0.720	0.719	0.173	0.184	0.795	0.707	0.810	0.717
			1.25	0.917	0.683	0.623	0.747	0.743	0.128	0.132	0.823	0.724	0.843	0.735
5		0.75	—	0.811	0.757	0.808	0.533	0.762	0.563	0.818	0.583	0.815	0.584	
		1	—	0.865	0.800	0.867	0.566*	0.803	0.563	0.875	0.607	0.872	0.611	
		1.25	—	0.892	0.820	0.896	0.580	0.814	0.481	0.904	0.617	0.899	0.622	
7		0.75	—	0.835	0.768	0.836	0.509	0.761	0.514	0.840	0.541	0.838	0.540	
		1	—	0.890	0.810	0.893	0.538	0.791	0.490	0.896	0.562	0.893	0.563	
		1.25	—	0.917	0.830	0.921	0.553	0.793	0.385	0.924	0.573	0.921	0.574	
750		3	0.75	0.820	0.614	0.574	0.668	0.675	0.242	0.273	0.739	0.673	0.749	0.681
			1	0.889	0.663	0.611	0.721*	0.721	0.175	0.200	0.795	0.707	0.813	0.719
			1.25	0.925	0.689	0.628	0.747	0.743	0.129	0.150	0.824	0.723	0.846	0.737
	5	0.75	—	0.816	0.761	0.809	0.531	0.784	0.566	0.818	0.583	0.816	0.586	
		1	—	0.869	0.804	0.867	0.563	0.829	0.608	0.875	0.606	0.873	0.611	
		1.25	—	0.896	0.825	0.896	0.578	0.846	0.627	0.904	0.616	0.901	0.622	
	7	0.75	—	0.840*	0.771	0.837	0.509	0.807	0.539	0.841	0.541	0.840	0.542	
		1	—	0.893	0.813	0.893	0.535	0.849	0.572	0.896	0.560	0.895	0.562	
		1.25	—	0.921	0.834	0.922	0.547	0.864	0.584	0.925	0.569	0.923	0.571	

Table 6: Standard deviations of the Q^2 metrics of the 500 samples for each experimental conditions, averaged over model and prediction method

n	i	l	EM	ML	PLSc		PLSPM	
					Mode A	Mode B	Mode A	Mode B
100	3	0.75	0.03520	0.02820	0.02490	0.07226	0.02424	0.02597
		1	0.02413	0.02389	0.01911	0.07513	0.01859	0.02069
		1.25	0.01692	0.02168	0.01612	0.08580	0.01583	0.01760
	5	0.75	0.03777	0.02597	0.02456	0.11873	0.02343	0.02448
		1	0.02377	0.02280	0.02023	0.15187	0.01927	0.02027
		1.25	0.01674	0.01947	0.01621	0.16472	0.01549	0.01675
	7	0.75	—	0.02575	0.02427	0.14189	0.02353	0.02523
		1	—	0.02167	0.01941	0.16580	0.01887	0.02065
		1.25	—	0.01915	0.01647	0.17096	0.01605	0.01771
250	3	0.75	0.01688	0.01536	0.01515	0.02160	0.01513	0.01541
		1	0.01175	0.01375	0.01185	0.03230	0.01180	0.01227
		1.25	0.00808	0.01212	0.00984	0.03620	0.00984	0.01031
	5	0.75	0.01653	0.01477	0.01515	0.02522	0.01450	0.01473
		1	0.01084	0.01220	0.01212	0.05728	0.01158	0.01178
		1.25	0.00774	0.01073	0.01024	0.11178	0.00981	0.01001
	7	0.75	—	0.01527	0.01525	0.03527	0.01479	0.01519
		1	—	0.01195	0.01199	0.06739	0.01195	0.01199
		1.25	—	0.01055	0.01021	0.11396	0.00994	0.01036
750	3	0.75	0.00869	0.00840	0.00849	0.01011	0.00851	0.00857
		1	0.00603	0.00688	0.00659	0.00925	0.00666	0.00669
		1.25	0.00423	0.00588	0.00531	0.00887	0.00541	0.00542
	5	0.75	0.00844	0.00828	0.00880	0.00956	0.00841	0.00846
		1	0.00564	0.00658	0.00689	0.00876	0.00660	0.00663
		1.25	0.00413	0.00570	0.00606	0.00995	0.00580	0.00585
	7	0.75	—	0.00811	0.00853	0.01043	0.00826	0.00833
		1	—	0.00656	0.00704	0.01091	0.00683	0.00686
		1.25	—	0.00524	0.00560	0.01419	0.00545	0.00553

Table 7: Mean Q^2 for model 1 with misspecification. All values significantly different from next lowest in row (paired sample t-test, $p=0.05$), except where noted by *

n	i	l	ML		PLSc (mode A)		PLSPM (mode A)	
			comm	red	comm	red	comm	red
100	3	0.75	0.734	0.733*	0.716	0.732	0.747	0.730
		1	0.790	0.788	0.775	0.792	0.804	0.791*
		1.25	0.815	0.812	0.804	0.821*	0.833	0.821
	5	0.75	0.797	0.770	0.806	0.756	0.817	0.755
		1	0.852	0.818	0.863	0.810	0.873	0.809
		1.25	0.878	0.840	0.893	0.837	0.902	0.837
	7	0.75	0.824	0.785	0.833	0.768	0.839	0.767
		1	0.880	0.832	0.890	0.821	0.895	0.820
		1.25	0.908	0.854	0.919	0.847	0.923	0.846
250	3	0.75	0.746	0.743	0.718	0.736	0.748	0.731
		1	0.801	0.797	0.777	0.795	0.805	0.792
		1.25	0.828	0.824	0.807	0.825*	0.834	0.823
	5	0.75	0.806	0.778	0.807	0.758	0.817	0.755
		1	0.860	0.826	0.866	0.813	0.875	0.811
		1.25	0.888	0.850	0.895	0.840	0.904	0.839
	7	0.75	0.833	0.793	0.835	0.770	0.839	0.769
		1	0.888	0.840	0.892	0.823	0.896	0.821
		1.25	0.917	0.863	0.921	0.849	0.924	0.848
750	3	0.75	0.752	0.749	0.720	0.738	0.748	0.733
		1	0.805	0.801	0.777	0.797	0.805	0.793
		1.25	0.833	0.828	0.807	0.827	0.834	0.824
	5	0.75	0.810	0.782	0.808	0.759	0.818	0.756
		1	0.865	0.830	0.866	0.813	0.875	0.811
		1.25	0.893	0.854	0.896	0.840	0.904	0.839
	7	0.75	0.837	0.797	0.836	0.771	0.839	0.769
		1	0.892	0.844	0.893	0.824	0.896	0.822
		1.25	0.920	0.867	0.922	0.850	0.925	0.849

Table 8: Mean Q^2 for model 3 with misspecification. All values significantly different from next lowest in row (paired sample t-test, $p=0.05$), except where noted by *

n	i	l	ML		PLSc (mode A)		PLSPM (mode A)	
			comm	red	comm	red	comm	red
100	3	0.75	0.598	0.561	0.666*	0.665	0.737	0.671
		1	0.647	0.598	0.720	0.712	0.794	0.708
		1.25	0.670	0.611	0.745	0.734	0.821	0.722
	5	0.75	0.801	0.748	0.806	0.539	0.816	0.586
		1	0.855	0.790	0.864	0.570	0.873	0.608
		1.25	0.881	0.809	0.893	0.585	0.902	0.619
	7	0.75	0.826	0.759	0.833	0.513	0.838	0.542
		1	0.880	0.800	0.890	0.542	0.894	0.563
		1.25	0.908	0.821	0.919	0.557	0.923	0.574
250	3	0.75	0.610	0.571	0.668	0.671	0.738	0.673
		1	0.657	0.606	0.721	0.717	0.795	0.707
		1.25	0.683	0.623	0.748	0.739	0.823	0.723
	5	0.75	0.811	0.756	0.807	0.532	0.817	0.582
		1	0.864	0.799	0.866	0.565	0.875	0.607
		1.25	0.891	0.820	0.895	0.579	0.903	0.616
	7	0.75	0.835	0.767	0.835	0.510	0.839	0.542
		1	0.890	0.810	0.893	0.537	0.896	0.561
		1.25	0.917	0.830	0.921	0.550	0.924	0.571
750	3	0.75	0.614	0.574	0.668	0.673	0.738	0.672
		1	0.662	0.610	0.721	0.719	0.795	0.707
		1.25	0.689	0.628	0.748	0.742	0.824	0.723
	5	0.75	0.815	0.761	0.809	0.532	0.818	0.583
		1	0.869	0.803	0.867	0.562	0.875	0.605
		1.25	0.896	0.824	0.896	0.577	0.904	0.616
	7	0.75	0.839	0.771	0.836	0.507	0.840	0.541
		1	0.893*	0.813	0.893	0.535	0.896	0.560
		1.25	0.921	0.834	0.922	0.547	0.925	0.569

Table 9: Mean Q^2 for model 3 with misspecification where the estimated model omits correct paths and adds wrong paths. All values significantly different from next lowest in row (paired sample t-test, $p=0.05$), except where noted by *

n	i	l	ML		PLSc (mode A)		PLSPM (mode A)	
			comm	red	comm	red	comm	red
100	3	0.75	0.593	0.456	0.659	0.565	0.737	0.584
		1	0.643	0.479	0.712	0.608	0.794	0.626
		1.25	0.671	0.500	0.740	0.629	0.822	0.646
	5	0.75	0.796	0.570*	0.802	0.546	0.815	0.565
		1	0.852	0.630	0.862	0.587	0.874	0.602
		1.25	0.878	0.635	0.891	0.604	0.902	0.618
	7	0.75	0.826	0.564	0.832	0.540	0.839	0.552
		1	0.879	0.632	0.888	0.576	0.895	0.585
		1.25	0.908	0.649	0.918	0.596	0.923	0.604
250	3	0.75	0.602	0.467	0.665	0.571	0.738	0.584
		1	0.655	0.485	0.718	0.615	0.795	0.627
		1.25	0.681	0.504	0.744	0.635	0.823	0.647
	5	0.75	0.807*	0.600	0.807	0.552	0.818	0.569
		1	0.862	0.645	0.865	0.591	0.875	0.605
		1.25	0.889	0.647	0.894	0.609	0.903	0.622
	7	0.75	0.833	0.589	0.834	0.541	0.839	0.553
		1	0.888	0.643	0.891	0.578	0.895	0.587
		1.25	0.917	0.651	0.920	0.597	0.924	0.605
750	3	0.75	0.609	0.478	0.668	0.576	0.739	0.586
		1	0.659	0.490	0.720	0.618	0.795	0.627
		1.25	0.687	0.501	0.746	0.637	0.824	0.647
	5	0.75	0.811	0.616	0.808	0.553	0.818	0.569
		1	0.866	0.652	0.867	0.593	0.876	0.607
		1.25	0.894	0.651	0.896	0.613	0.904	0.625
	7	0.75	0.837	0.610	0.836	0.543	0.840	0.554
		1	0.892	0.649	0.893	0.580	0.896	0.588
		1.25	0.920	0.659	0.922	0.599	0.925	0.606

Table 10: Percentage difference in Q^2 for underspecified model.

n	i	l	ML		PLSc (mode A)		PLSPM (mode A)	
			comm	red	comm	red	comm	red
100	3	0.75	-0.5454	-6.9181	-0.4359	-1.1827	-0.0405	-2.4995
		1	-0.3002	-7.2048	-0.4352	-0.9588	-0.0395	-2.3495
		1.25	-0.1925	-7.3944	-0.4380	-0.7166	-0.0401	-2.1405
	5	0.75	-0.7035	-7.6449	-0.0329	-6.9010	-0.0018	-5.8420
		1	-0.6143	-7.7749	-0.0263	-6.4748	-0.0019	-5.7185
		1.25	-0.5862	-7.8718	-0.0220	-6.2336	-0.0017	-5.6155
	7	0.75	-0.3592	-7.9519	-0.0152	-7.8174	-0.0008	-7.1727
		1	-0.2850	-8.0104	-0.0122	-7.3030	-0.0006	-6.9295
		1.25	-0.2518	-8.0382	-0.0087	-7.0410	-0.0007	-6.7832
250	3	0.75	-0.5495	-6.8629	-0.4187	-1.0591	-0.0345	-2.4452
		1	-0.2913	-7.1525	-0.4135	-0.7972	-0.0333	-2.2601
		1.25	-0.1838	-7.3875	-0.4055	-0.6698	-0.0333	-2.1842
	5	0.75	-0.7365	-7.5547	-0.0159	-7.0516	-0.0013	-5.8847
		1	-0.6607	-7.7071	-0.0144	-6.6668	-0.0010	-5.8018
		1.25	-0.6306	-7.7990	-0.0131	-6.4707	-0.0010	-5.7504
	7	0.75	-0.3753	-7.8821	-0.0065	-8.0895	-0.0003	-7.3131
		1	-0.3075	-7.9629	-0.0038	-7.5028	-0.0004	-7.0072
		1.25	-0.2746	-7.9322	-0.0040	-7.2074	-0.0003	-6.8320
750	3	0.75	-0.5554	-6.9911	-0.4108	-1.1713	-0.0312	-2.5996
		1	-0.2955	-7.2210	-0.4085	-0.8676	-0.0303	-2.3692
		1.25	-0.1818	-7.4301	-0.4098	-0.7554	-0.0289	-2.2932
	5	0.75	-0.7545	-7.5173	-0.0103	-7.1387	-0.0008	-5.9150
		1	-0.6804	-7.6349	-0.0099	-6.6597	-0.0007	-5.7557
		1.25	-0.6482	-7.7067	-0.0094	-6.4739	-0.0007	-5.7092
	7	0.75	-0.3856	-7.7792	-0.0029	-8.0728	-0.0001	-7.2451
		1	-0.3159	-7.9146	-0.0024	-7.6029	-0.0001	-7.0503
		1.25	-0.2855	-7.9488	-0.0021	-7.3403	-0.0001	-6.9123

Table 11: Experimental Design Factors for Study 2

Factor loadings (unstandardized)	l	1
Indicator covariance	c	0, 0.1, 0.4
Error variance on formative constructs	e	0, 0.1, 0.4
Estimation method		PLSPM, ML, LM
Outer modes (for PLSPM)		Mode A, Mode B, Mixed
Inner schemes (for PLSPM)		Centroid, Path, Factor

Table 12: Mean RMSE for model 1 with 10-fold cross-validation. Non-significance (paired-sample t-test, $p < 0.05$) to next lowest value with equal n, i, e, c indicated by *

n	i	e	PLSPM-A			PLSPM-B			LM			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	0.419	0.406	0.396	0.419	0.416	0.417	0.421	0.419	0.419	0.473	0.469	0.475
		0.1	0.508	0.500	0.489	0.509	0.510	0.508	0.519	0.519	0.518	0.564	0.565	0.564
		0.4	0.714	0.708	0.700	0.715	0.716	0.717	0.738	0.740	0.740	0.801	0.806	0.801
	5	0	0.442	0.414	0.396	0.441	0.441	0.440	0.436	0.436	0.436	0.478	0.474	0.477
		0.1	0.530	0.506	0.49	0.531	0.531	0.530	0.540	0.541	0.541	0.569	0.572	0.571
		0.4	0.733	0.710	0.700	0.738	0.732	0.736	0.772	0.768	0.771	0.823	0.814	0.817
	7	0	0.466	0.417	0.396	*0.466	0.464	0.465	0.456	0.456	0.456	0.479	0.480	0.479
		0.1	0.550	0.508	0.489	0.553	0.551	0.550	0.566	0.564	0.564	0.578	0.584	*0.609
		0.4	0.747	0.720	0.701	0.756	0.761	0.757	0.806	0.810	0.806	0.831	0.838	0.870
250	3	0	0.404	0.400	0.396	0.404	0.403	0.404	0.404	0.404	0.405	0.426	0.425	0.426
		0.1	*0.498	0.493	0.491	0.498	0.497	0.498	0.501	0.501	0.502	0.525	0.524	0.525
		0.4	0.708	0.706	0.702	0.708	0.709	0.708	0.717	0.718	0.717	0.772	0.773	0.771
	5	0	0.415	0.403	0.396	0.414	0.413	0.413	0.411	0.411	0.411	0.428	0.428	0.427
		0.1	0.506	0.496	0.491	0.506	0.506	0.506	0.509	0.509	0.509	0.528	0.527	0.528
		0.4	0.716	0.706	0.701	0.717	0.714	0.714	0.729	0.727	0.727	0.778	0.775	0.776
	7	0	0.423	0.404	0.396	0.422	0.421	0.422	0.417	0.417	0.417	0.428	0.429	0.429
		0.1	0.515	0.497	0.492	0.515	0.513	0.515	0.517	0.516	0.518	0.530	0.503	0.531
		0.4	0.723	0.707	0.701	0.724	0.721	0.722	0.740	0.737	0.738	0.784	0.780	0.781
750	3	0	0.399	0.397	0.396	0.399	0.399	0.399	*0.399	0.399	0.399	0.411	0.411	0.411
		0.1	0.493	0.492	0.491	0.493	0.493	0.493	0.494	0.494	0.494	0.513	0.513	0.513
		0.4	0.705	0.704	0.702	*0.705	0.705	0.704	0.708	0.708	0.707	0.764	0.764	0.763
	5	0	0.402	0.398	0.396	0.401	0.401	0.401	0.401	0.401	0.401	0.412	0.411	0.412
		0.1	0.496	0.493	0.491	0.496	0.496	0.496	0.497	0.497	0.497	0.514	0.514	0.514
		0.4	0.707	0.705	0.702	0.707	0.708	0.707	0.711	0.711	0.710	0.765	0.766	0.765
	7	0	0.405	0.398	0.396	0.404	0.404	0.404	0.403	0.402	0.403	0.412	0.412	0.412
		0.1	0.499	0.493	0.491	0.499	0.499	0.498	0.500	0.499	0.499	0.515	0.515	0.515
		0.4	*0.710	0.705	0.702	0.710	0.709	0.709	0.715	0.714	0.714	0.768	0.767	0.766

Table 13: Mean RMSE for model 2 with 10-fold cross-validation. Non-significance (paired-sample t-test, $p < 0.05$) to next lowest value with equal n, i, e, c indicated by *

n	i	e	PLSPM-A			PLSPM-B			LM			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	0.461	0.458	0.458	0.460	0.459	0.459	0.472	0.472	0.472	0.530	0.531	0.529
		0.1	0.539	0.538	0.534	0.539	0.539	0.537	0.552	0.552	0.551	0.620	*0.672	0.607
		0.4	0.723	0.720	0.718	0.724	0.723	0.723	0.738	0.737	0.738	0.821	0.818	0.816
	5	0	0.463	0.461	0.459	0.461	0.461	0.461	0.478	0.478	0.479	0.528	0.531	0.527
		0.1	0.542	0.538	0.535	0.540	0.541	0.540	0.558	0.559	0.558	0.609	0.617	0.608
		0.4	0.730	0.722	0.719	0.731	0.729	0.730	0.748	0.747	0.747	0.828	0.822	0.825
	7	0	0.469	0.461	0.458	0.463	0.462	0.461	0.485	0.484	0.484	0.548	0.527	0.529
		0.1	0.548	0.538	0.535	0.544	0.543	0.543	0.565	0.564	0.565	0.61	0.634	0.609
		0.4	0.735	0.722	0.716	*0.735	0.735	0.733	0.755	0.755	0.753	0.832	0.832	0.825
250	3	0	0.461	0.461	0.461	0.461	0.461	0.461	0.466	0.466	0.466	0.491	0.491	0.491
		0.1	0.539	0.539	0.539	0.539	0.540	0.540	0.544	0.545	0.545	0.576	0.576	0.576
		0.4	0.725	0.724	0.724	*0.725	0.725	0.726	0.731	0.731	0.731	0.796	0.796	0.797
	5	0	0.463	0.462	0.461	0.462	0.462	0.461	0.468	0.469	0.468	0.491	0.491	0.492
		0.1	0.542	0.539	0.538	0.541	0.540	0.540	0.548	0.547	0.547	0.576	0.576	0.576
		0.4	0.729	0.726	0.723	*0.729	0.729	0.727	0.735	0.736	0.734	0.800	0.802	0.800
	7	0	0.464	0.462	0.461	0.462	0.462	0.462	0.471	0.471	0.471	0.492	0.492	0.493
		0.1	0.543	0.540	0.538	0.542	0.541	0.541	0.550	0.550	0.549	0.578	0.577	0.577
		0.4	0.730	0.726	0.724	0.730	0.730	0.730	0.738	0.738	0.738	0.803	0.804	0.804
750	3	0	0.462	0.462	0.462	0.462	0.462	0.462	0.464	0.464	0.464	0.481	0.481	0.481
		0.1	0.540	0.540	0.540	0.540	0.540	0.540	0.542	0.542	0.542	0.567	0.567	0.567
		0.4	0.726	0.727	0.725	*0.726	0.727	0.726	0.728	0.729	0.728	0.793	0.794	0.793
	5	0	0.462	0.462	0.462	0.462	0.462	0.462	0.464	0.464	0.464	0.481	0.481	0.481
		0.1	0.541	0.540	0.540	0.541	0.541	0.540	0.543	0.543	0.543	0.568	0.568	0.567
		0.4	0.728	0.726	0.725	0.728	0.726	0.726	0.730	0.729	0.728	0.794	0.793	0.793
	7	0	0.463	0.462	0.462	0.462	0.462	0.462	0.465	0.465	0.465	0.481	0.481	0.481
		0.1	0.542	0.540	0.540	0.541	0.541	0.541	0.544	0.543	0.544	0.568	0.568	0.568
		0.4	0.728	0.727	0.726	0.728	0.728	0.728	0.731	0.731	0.731	0.795	0.795	0.795

Table 14: Mean RMSE for model 3 with 10-fold cross-validation. Non-significance (paired-sample t-test, $p < 0.05$) to next lowest value with equal n, i, e, c indicated by *

n	i	e	PLSPM-A			PLSPM-B			LM			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	0.473	0.464	0.455	0.472	0.472	0.472	0.483	0.483	0.482	0.543	0.533	0.532
		0.1	0.549	0.542	0.534	0.549	0.549	0.549	0.565	0.566	0.566	0.606	0.617	0.610
		0.4	0.735	0.728	0.721	0.735	0.735	0.736	0.764	0.763	0.764	0.829	0.845	0.834
	5	0	0.491	0.469	0.455	0.489	0.489	0.490	0.502	0.502	0.501	0.533	0.532	0.530
		0.1	0.569	0.547	0.535	0.568	0.566	0.567	0.592	0.589	0.590	0.621	0.614	0.622
		0.4	0.751	0.734	0.721	0.752	0.753	0.750	0.796	0.797	0.794	0.846	0.842	0.844
	7	0	0.512	0.473	0.454	0.507	0.508	0.507	0.525	0.525	0.524	0.535	0.538	0.539
		0.1	0.588	0.549	0.533	0.585	0.582	0.582	0.616	0.615	0.616	0.62	0.618	0.626
		0.4	0.761	0.733	0.719	0.763	0.764	0.764	0.828	0.828	0.827	0.868	0.852	0.888
250	3	0	0.462	0.459	0.455	0.462	0.462	0.462	0.465	0.465	0.465	0.486	0.487	0.487
		0.1	0.542	0.537	0.535	0.542	0.54	0.541	0.548	0.546	0.547	0.574	0.572	0.573
		0.4	0.729	0.725	0.722	*0.729	0.728	0.728	0.739	0.739	0.738	0.80	0.799	0.798
	5	0	0.470	0.460	0.455	0.469	0.468	0.468	0.472	0.472	0.472	0.488	0.488	0.488
		0.1	0.550	0.540	0.535	0.549	0.547	0.547	0.556	0.555	0.555	0.577	0.576	0.576
		0.4	0.735	0.726	0.724	*0.735	0.733	0.736	0.750	0.748	0.751	0.805	0.803	0.806
	7	0	0.478	0.461	0.455	0.476	0.475	0.475	0.479	0.479	0.479	0.489	0.489	0.488
		0.1	0.554	0.541	0.535	0.553	0.554	0.554	0.562	0.563	0.563	0.577	0.577	0.578
		0.4	0.74	0.727	0.722	0.740	0.740	0.740	0.760	0.760	0.760	0.810	0.809	0.809
750	3	0	0.458	0.457	0.455	0.457	0.458	0.458	0.458	0.458	0.458	0.474	0.475	0.475
		0.1	0.538	0.536	0.535	0.538	0.537	0.537	0.539	0.539	0.539	0.563	0.562	0.563
		0.4	0.725	0.725	0.723	0.725	0.726	0.725	0.729	0.729	0.729	0.791	0.791	0.791
	5	0	0.460	0.457	0.455	0.459	0.459	0.460	0.461	0.460	0.461	0.475	0.475	0.475
		0.1	0.539	0.537	0.535	0.539	0.539	0.539	0.541	0.541	0.542	0.563	0.564	0.564
		0.4	0.728	0.726	0.723	0.728	0.728	0.727	0.732	0.733	0.732	0.793	0.794	0.793
	7	0	0.462	0.457	0.455	0.462	0.461	0.462	0.463	0.462	0.463	0.475	0.475	0.475
		0.1	0.542	0.537	0.535	0.541	0.541	0.541	0.544	0.544	0.544	0.564	0.564	0.564
		0.4	0.730	0.726	0.724	0.730	0.730	0.730	0.736	0.736	0.736	0.795	0.795	0.795

Table 15: Mean standard deviations of RMSE across all models.

n	i	e	PLSPM-A			PLSPM-B			LM			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	0.013	0.011	0.01	0.013	0.013	0.013	0.012	0.011	0.011	0.188	0.088	0.068
		0.1	0.02	0.019	0.018	0.02	0.02	0.019	0.02	0.02	0.019	0.137	0.529	0.066
		0.4	0.033	0.031	0.031	0.033	0.032	0.032	0.034	0.033	0.033	0.074	0.158	0.059
	5	0	0.015	0.011	0.009	0.015	0.015	0.015	0.011	0.011	0.011	0.067	0.051	0.04
		0.1	0.021	0.019	0.017	0.022	0.022	0.022	0.021	0.021	0.02	0.054	0.068	0.08
		0.4	0.033	0.032	0.031	0.034	0.034	0.034	0.036	0.036	0.036	0.085	0.058	0.06
	7	0	0.016	0.011	0.009	0.016	0.017	0.017	0.012	0.012	0.011	0.183	0.067	0.062
		0.1	0.023	0.019	0.018	0.024	0.023	0.024	0.022	0.022	0.022	0.042	0.243	0.294
		0.4	0.034	0.031	0.029	0.035	0.035	0.035	0.038	0.038	0.038	0.139	0.062	0.463
250	3	0	0.007	0.007	0.006	0.007	0.007	0.007	0.006	0.006	0.006	0.01	0.011	0.011
		0.1	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.014	0.014	0.014
		0.4	0.019	0.019	0.018	0.019	0.019	0.019	0.019	0.02	0.019	0.026	0.025	0.026
	5	0	0.007	0.006	0.005	0.007	0.007	0.007	0.006	0.006	0.006	0.01	0.01	0.01
		0.1	0.012	0.011	0.011	0.012	0.012	0.012	0.012	0.011	0.012	0.015	0.015	0.015
		0.4	0.019	0.019	0.018	0.019	0.019	0.019	0.02	0.02	0.019	0.025	0.026	0.026
	7	0	0.007	0.006	0.005	0.007	0.007	0.008	0.006	0.006	0.006	0.01	0.01	0.01
		0.1	0.012	0.011	0.01	0.012	0.012	0.012	0.012	0.011	0.011	0.014	0.014	0.014
		0.4	0.019	0.019	0.018	0.019	0.02	0.019	0.02	0.02	0.019	0.026	0.026	0.025
750	3	0	0.004	0.003	0.003	0.004	0.004	0.004	0.004	0.003	0.003	0.005	0.005	0.005
		0.1	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.008	0.007	0.008
		0.4	0.011	0.01	0.01	0.011	0.01	0.01	0.011	0.01	0.01	0.014	0.014	0.014
	5	0	0.004	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.004
		0.1	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.007
		0.4	0.011	0.011	0.01	0.011	0.011	0.01	0.011	0.011	0.01	0.014	0.014	0.014
	7	0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.004	0.004
		0.1	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.007
		0.4	0.011	0.01	0.01	0.011	0.01	0.01	0.011	0.01	0.01	0.014	0.013	0.014

Table 16: RMSE for the misspecified model 1 with 10-fold cross-validation. All values with equal n, i, e, c are significantly different (paired-sample t-test, $p < 0.05$).

n	i	e	PLSPM-A			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	0.429	0.415	0.400	0.469	0.475	0.475
		0.1	0.521	0.506	0.494	0.570	0.564	0.568
		0.4	0.727	0.716	0.702	0.815	0.826	0.804
	5	0	0.462	0.428	0.399	0.474	0.480	0.475
		0.1	0.547	0.516	0.493	0.578	0.581	0.574
		0.4	0.747	0.720	0.701	0.844	0.820	0.820
	7	0	0.494	0.431	0.399	0.479	0.480	0.493
		0.1	0.574	0.521	0.494	0.582	0.601	0.580
		0.4	0.757	0.727	0.706	0.827	0.891	0.842
250	3	0	0.412	0.405	0.397	0.427	0.427	0.426
		0.1	0.504	0.498	0.492	0.526	0.526	0.526
		0.4	0.713	0.710	0.703	0.774	0.775	0.773
	5	0	0.426	0.410	0.397	0.428	0.428	0.427
		0.1	0.517	0.502	0.493	0.528	0.528	0.530
		0.4	0.722	0.711	0.703	0.777	0.778	0.778
	7	0	0.441	0.412	0.397	0.430	0.429	0.429
		0.1	0.529	0.504	0.493	0.530	0.530	0.532
		0.4	0.734	0.714	0.702	0.785	0.785	0.782
750	3	0	0.402	0.399	0.397	0.411	0.411	0.412
		0.1	0.496	0.494	0.491	0.513	0.513	0.513
		0.4	0.706	0.705	0.703	0.763	0.764	0.764
	5	0	0.408	0.401	0.396	0.412	0.411	0.411
		0.1	0.501	0.495	0.491	0.515	0.514	0.514
		0.4	0.711	0.706	0.703	0.766	0.766	0.766
	7	0	0.413	0.401	0.397	0.412	0.412	0.413
		0.1	0.505	0.495	0.492	0.515	0.515	0.516
		0.4	0.714	0.707	0.703	0.768	0.768	0.767

Table 17: RMSE for the misspecified model 3 with 10-fold cross-validation. Non-significance (paired-sample t-test, $p < 0.05$) to next lowest value with equal n, i, e, c indicated by *

n	i	e	PLSPM-A			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	0.488	0.475	0.458	0.527	0.556	0.531
		0.1	0.564	0.549	0.537	0.612	0.612	0.609
		0.4	0.746	0.737	0.722	0.839	0.831	0.828
	5	0	0.519	0.485	0.458	0.533	0.537	0.538
		0.1	0.591	0.559	0.536	0.619	0.621	0.621
		0.4	0.764	0.741	0.724	0.849	0.844	0.849
	7	0	0.545	0.489	0.458	*0.626	0.543	0.541
		0.1	0.615	0.566	0.535	0.622	0.628	0.632
		0.4	0.783	0.745	0.722	0.862	0.860	0.863
250	3	0	0.470	0.463	0.457	0.487	0.487	0.487
		0.1	0.550	0.543	0.537	0.575	0.574	0.575
		0.4	0.732	0.730	0.724	0.799	0.801	0.800
	5	0	0.485	0.468	0.456	0.488	0.488	0.487
		0.1	0.563	0.547	0.536	0.577	0.576	0.577
		0.4	0.744	0.733	0.724	0.805	0.806	0.806
	7	0	0.499	0.471	0.455	0.490	0.490	0.489
		0.1	0.575	0.549	0.536	0.578	0.578	0.578
		0.4	0.754	0.734	0.723	0.812	0.811	0.810
750	3	0	0.462	0.458	0.456	0.475	0.474	0.475
		0.1	0.540	0.538	0.536	0.563	0.562	0.563
		0.4	0.728	0.726	0.725	0.792	0.791	0.792
	5	0	0.467	0.460	0.455	0.475	0.475	0.475
		0.1	0.546	0.539	0.535	0.564	0.564	0.563
		0.4	0.732	0.727	0.724	0.793	0.794	0.794
	7	0	0.472	0.461	0.455	0.475	0.475	0.475
		0.1	0.550	0.540	0.535	0.564	0.565	0.564
		0.4	0.735	0.727	0.724	0.795	0.794	0.795

Table 18: RMSE for the misspecified model 3 (wrong model) with 10-fold cross-validation.

All values with equal n, i, e, c are significantly different (paired-sample t-test, $p < 0.05$).

n	i	e	PLPM-A			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	0.659	0.654	0.647	0.782	0.818	0.773
		0.1	0.699	0.696	0.693	0.832	0.834	0.829
		0.4	0.814	0.808	0.804	0.969	0.964	0.968
	5	0	0.669	0.657	0.649	0.79	0.779	0.779
		0.1	0.710	0.699	0.693	0.829	0.826	0.844
		0.4	0.825	0.811	0.804	0.985	0.976	0.978
	7	0	0.681	0.659	0.650	0.785	0.782	0.785
		0.1	0.720	0.701	0.692	0.832	0.835	0.840
		0.4	0.830	0.814	0.806	0.987	0.991	0.994
250	3	0	0.654	0.653	0.652	0.753	0.754	0.754
		0.1	0.696	0.695	0.694	0.800	0.799	0.802
		0.4	0.812	0.810	0.809	0.946	0.948	0.949
	5	0	0.658	0.654	0.651	0.753	0.754	0.754
		0.1	0.702	0.697	0.694	0.804	0.803	0.804
		0.4	0.815	0.812	0.807	0.949	0.952	0.949
	7	0	0.662	0.656	0.653	0.755	0.757	0.758
		0.1	0.706	0.698	0.693	0.805	0.806	0.804
		0.4	0.820	0.812	0.806	0.956	0.955	0.952
750	3	0	0.654	0.653	0.653	0.750	0.749	0.749
		0.1	0.696	0.696	0.695	0.797	0.798	0.797
		0.4	0.811	0.810	0.809	0.943	0.942	0.942
	5	0	0.655	0.653	0.653	0.750	0.749	0.751
		0.1	0.698	0.696	0.695	0.798	0.799	0.798
		0.4	0.812	0.810	0.810	0.944	0.943	0.944
	7	0	0.656	0.653	0.653	0.749	0.750	0.751
		0.1	0.699	0.696	0.695	0.800	0.799	0.799
		0.4	0.813	0.810	0.810	0.944	0.945	0.945

Table 19: Percentage difference in RMSE for underspecified model.

n	i	e	PLSPM-A			ML		
			c=0	c=0.1	c=0.4	c=0	c=0.1	c=0.4
100	3	0	25.0	24.7	24.6	16.0	15.1	14.1
		0.1	14.4	14.0	13.5	10.3	11.0	10.6
		0.4	4.2	4.0	3.8	4.0	3.8	4.4
	5	0	24.8	24.8	24.5	16.3	16.4	17.2
		0.1	14.7	14.0	13.5	10.5	8.9	9.5
		0.4	4.5	4.0	3.7	4.3.0	3.9	5.4
	7	0	24.6	24.7	24.4	16.6	16.8	16.6
		0.1	14.9	14.1	13.3	13.5	10.8	10.3
		0.4	4.7	4.1	3.6	3.6	4.0	4.4
250	3	0	24.8	24.7	24.5	20.4	20.4	20.3
		0.1	13.8	13.8	13.5	12.2	12.3	12.2
		0.4	3.9	3.8	3.7	3.9	3.9	3.9
	5	0	25.0	24.7	24.6	21.3	21.3	21.3
		0.1	14.0	13.5	13.3	12.6	12.5	12.5
		0.4	4.0	3.8	3.6	4.0	4.0	4.0
	7	0	24.6	24.7	24.6	21.6	21.6	21.6
		0.1	14.1	13.5	13.3	12.7	12.8	12.8
		0.4	4.1	3.7	3.6	4.1	4.1	4.1
750	3	0	24.8	24.7	24.6	22.4	22.4	22.3
		0.1	13.5	13.5	13.4	12.8	12.8	12.8
		0.4	3.8	3.7	3.7	3.9	3.9	3.9
	5	0	24.8	24.6	24.6	23.3	23.2	23.2
		0.1	13.5	13.4	13.3	13.2	13.1	13.2
		0.4	3.7	3.6	3.6	3.9	3.9	4.0
	7	0	24.7	24.7	24.5	23.6	23.6	23.5
		0.1	13.6	13.3	13.2	13.4	13.3	13.3
		0.4	3.7	3.6	3.5	3.9	3.9	3.9