Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

STRUCTURAL EQUATION MODELING FOR SIMULATION METAMODELING

Kai G. Mertens
Iris Lorscheid
Matthias Meyer

Institute of Management Control and Accounting
Hamburg University of Technology
Am Schwarzenberg-Campus 4, 21073
Hamburg, GERMANY

ABSTRACT

The analysis of the behavior of simulation models and the subsequent communication of their results are critical but often neglected activities in simulation modeling. To overcome this issue, this paper proposes an integrated metamodeling approach based on structural equation modeling using the partial least squares algorithm. The suggested method integrates both a priori information from the conceptual model and the simulation data output. Based on this, we estimate and evaluate the core relationships and their predictive capabilities. The resulting structural equation metamodel exposes structures in the behavior of simulation models and allows for their better communication. The link to theory via the conceptual model considerably increases understanding compared with other metamodeling approaches.

1 INTRODUCTION

Both the analysis of the behavior of simulation models and the subsequent communication of their results are two important but often neglected activities in simulation modeling (Tolk 2013). Over time, several authors have stressed the importance of these two activities and suggested a number of approaches to conduct them more economically and effectively (Barton 2013; Sturrock 2014; Yilmaz et al. 2014). Typically, first-generation statistical techniques such as low-order polynomial regression metamodels are incorporated into these approaches. These metamodel techniques are suitable to provide insights into the effects of different input variables and help identify the main drivers of simulation results.

Still, these approaches face some challenges. While direct cause/effect relationships and interaction effects may be identified by using these approaches, more complex relationships between variables may be hidden. Current approaches do not incorporate the structure, strength, or direction of complex relationships between variables, such as hidden mediator effects in a more complex causal network that may include several layers of variables (Barton 2013, Sargent 1991). In particular, in agent-based models complex behavior can lead to several structural layers of effects, which are not covered explicitly by existing analytical approaches (Tolk et al. 2013). A related challenge is the analysis of a high number of variables and appropriate communication of their relationships and effects. Finally, low-order polynomial regression metamodels including analysis of variance (ANOVA) can lead to distorted results. Statistical characteristics about the data set – such as normally, independently, and identically distributed variables (NIID) – are defined as preconditions for their application and evaluation. However, these assumptions are often violated by simulation data.

This paper presents an approach to overcome some of the described challenges. To this end, we submit structural equation modeling (SEM) with its statistical algorithms to use both the a priori information on the conceptual simulation model and the data produced by the simulation model to estimate a structural equation metamodel. SEM is a second-generation statistical technique that provides a way of representing the simulation model representation that enables a semantic exposition of its structures over several layers and allows one to reduce complexity by aggregating and abstracting variables to constructs. Partial least squares (PLS-) SEM uses this information and the simulation data output in considering its statistical characteristics to estimate the parameter values (Hair et al. 2014). The resulting metamodel and its parameters can be assessed by several criteria to conclude its predictive capability and fit. The link between the conceptual model and theory may considerably increase understanding compared with other metamodeling approaches. Overall, the suggested structural equation metamodeling approach may foster the development of the analytical expositions of simulation models while keeping them communicable and understandable.

This paper proceeds as follows. In the next section, we introduce SEM, with an emphasis on PLS. First, we present its main characteristics and requirements, and then we discuss PLS-SEM as a suitable approach to manage the simulation data output characteristics and outline the different steps in deriving a metamodel. In Section 3, we illustrate the method with an example. First, we introduce the simulation model used and analyze it by adopting a regression technique. Then, we derive a metamodel based on the same simulation data and our conceptual model and discuss the benefits of the chosen approach. In Section 4, we provide a brief summary and conclusion.

2 METHODOLOGY

We propose that SEM be viewed as a metamodel technique. Generally, metamodels are input/output transformations that are inferred based on the simulation model (Kleijnen and Sargent 2000). However, SEM metamodels may also include the structure of the simulation model as information beyond the data. Thus, these models emphasize the structural aspect of metamodels.

All metamodels own specific key objectives such as understanding, prediction, optimization, and validation (Kleijnen 2014). Structural equation metamodels focus primarily on understanding. In this regard, they can be compared with the often used polynomial regression metamodels (Kleijnen and Sargent 2000). These approaches implicitly assume a direct cause/effect structure between the independent and dependent variables. Hence, the resulting metamodel embodies a simple structure of the simulation model (Law 2014b). However, this typically only allows one layer of effects. This characterizes first-generation statistical techniques (Gerow et al. 2008). SEM, as a second-generation statistical technique, can overcome this restriction by modeling mathematically complex structures.

2.1 PLS-SEM

SEM combines a conceptual semantic presentation of a model with a set of statistical techniques. Given the underlying research question, hypotheses are formulated a priori and establish a link to the theory. SEM visualizes these relationships between the entities to be investigated through its semantics. The corresponding measures are assigned to their entities to represent their meaning (Byrne 1998). The resulting structural equation model is estimated by means of a statistical algorithm. Thus, this methodology combines two different sources of information: (I) theory at the conceptual level and (II) data at the analytical level.

The estimation procedures of structural equation models fall into two groups (Hair et al. 2014): covariance-based algorithms, with an emphasis on explanation and parameter accuracy, and variance-based algorithms, with a distinct focus on the prediction and maximization of the explained variance. We focus on variance-based algorithm PLS. Its capability to reduce measurement errors to maximize the explained variance for a certain response is congruent with the aim of a simulation data analysis (Law 2014a). Furthermore, its emerging predictive capability is suitable for metamodeling, which will support the metamodel's plausibility.

The PLS algorithm was first presented by Wold (1982) and is a multiple ordinary least squares (OLS) regression technique for estimating composite factor models. As one can see from Figure 1, the SEM method consists of two different systems concerning the measurement and structural model. The structural model specifies the relationships between constructs and the measurement model describes the relationships between the construct and its corresponding variables. The constructs can also be called proxies, composites, or blocks and can interpreted as underlying factors produced by their corresponding measurement models (Falk and Miller 1992). PLS-SEM iteratively approximates the estimates of the measurement and structural parameters (Fornell and Bookstein 1982; Lohmüller 1989).

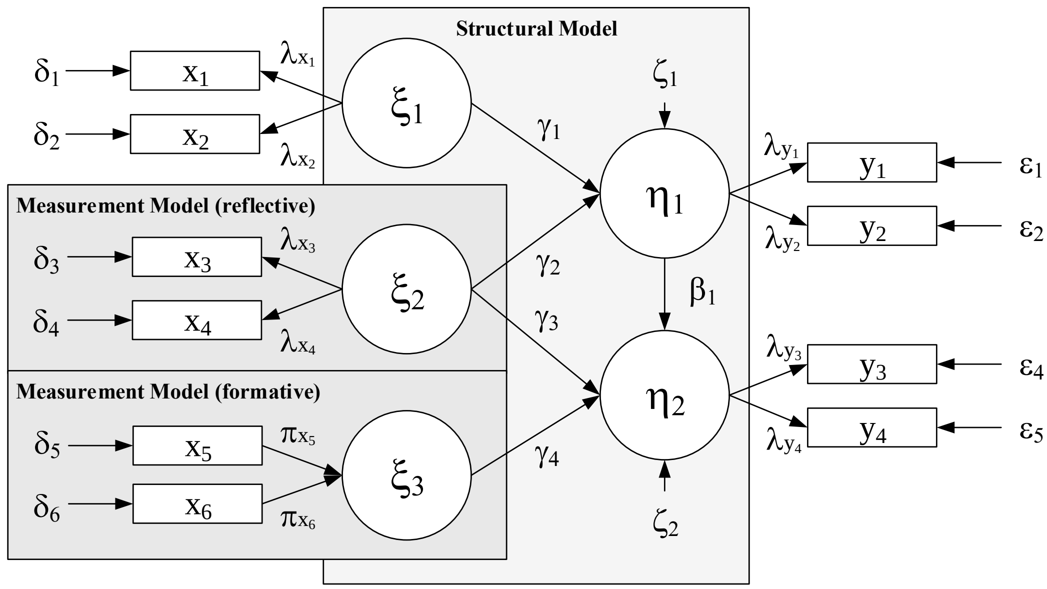

Figure 1: Structural Equation Model.

This structural model encompasses the constructs, which are either independent $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3 \dots \xi_n)$ or dependent $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots \eta_m)$. In particular, η_1 is determined as an intermediate construct. The structural model can easily be enhanced for more structural layers of effects and can be generally expressed by

$$E(\boldsymbol{\eta} \mid \boldsymbol{\eta}, \boldsymbol{\xi}) = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{1}$$

where $\mathbf{B}(m \times m)$ represents a matrix of the coefficient parameters for the dependent and intermediate constructs, and $\boldsymbol{\Gamma}(m \times n)$ is a matrix of the coefficient parameters for independent constructs $\boldsymbol{\xi}$. The terms $\boldsymbol{\zeta} = (\zeta_1, \zeta_2 \dots \zeta_{n\,x\,1})$ are residual vectors within the structural model and represent $\boldsymbol{\zeta} = \boldsymbol{\eta} - E(\boldsymbol{\eta})$. As noted earlier, the constructs are approximated by independent $\mathbf{x} = (x_1, x_2, \dots x_q)$ and dependent $\mathbf{y} = (y_1, y_2, \dots y_p)$ variables. There are two main measurement model types: formative and reflective (Hair et al. 2014). This relationship refers to the mathematical and conceptual link between the construct and its measure. The reflective measurement model can be mathematically described for independent, exogenous constructs by

$$\mathbf{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \qquad \sim E(\boldsymbol{\delta}) = 0 \tag{2}$$

and for dependent endogenous constructs by

$$\mathbf{y} = \mathbf{\Lambda}_{y}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \sim E(\boldsymbol{\varepsilon}) = 0 \tag{3}$$

$\mathbf{\Lambda}_{y}\,(p \times m)$ and $\mathbf{\Lambda}_{x}\,(q \times n)$ are regression matrices that are also called loadings, and $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are residual vectors. Thus, a reflective construct captures the common variance of its indicators. Furthermore, reflective measured constructs can be interpreted as abstractions, because the underlying construct represents the indicators.

The second model type is the formative measure, which is for independent constructs

$$\boldsymbol{\xi} = \mathbf{\Pi}_{\xi}\mathbf{x} \tag{4}$$

and for dependent variables

$$\boldsymbol{\eta} = \mathbf{\Pi}_{\eta}\mathbf{y}. \tag{5}$$

$\mathbf{\Pi}_{\eta}\,(p \times m)$ and $\mathbf{\Pi}_{\xi}\,(q \times n)$ represent the regression matrices that contain the weights. Hence, this measurement can be viewed as an aggregation of the variables, because the construct is caused by the indicators.

Accordingly, the estimation procedures can be roughly classified into three stages. First, the construct scores are approximated by the weights and loadings of their indicators. At the beginning of this procedure, initial arbitrary values are used. Then, the paths are approximated by the proxies that enable, simultaneously, an approximation of the scores by using the structural model. Based on these scores, the measurement model is estimated to deliver the parameter estimates for the first step. These stages are repeated iteratively until convergence occurs (I). The last steps are a final estimation of the path coefficients, weights, loadings (II), and location parameters (III) (Lohmüller 1989). Hence, PLS-SEM uses information from the structural model as well as data from the measurement model for its estimation. This property supports our aim of linking the conceptual and mathematical worlds to form a structural equation metamodel.

2.2 Suitability for Simulation

This section addresses PLS-SEM’s appropriateness for the analysis of simulated systems. Simulation data have specific requirements in comparison with the analysis of empirical data and models. In this respect, we will discuss the requirements of PLS for simulation models and its specific data characteristics.

First, simulation models often have complex internal structures such as high-order effects and hierarchical mechanisms. Basically, PLS-SEM covers these characteristics by its capability to model mediation, interaction, and non-linear effects. This is a valuable advantage in comparison with other regression metamodels, for which, for instance, mediation and several layers are quite challenging to detect and analyze (Lowry 2014).

Second, simulation models often consist of large numbers of variables and variable levels (Sanchez and Wan 2012). PLS-SEM is well known owing to its capability to analyze high-dimensional data in fuzzy environments (Sarstedt et al. 2014). The measurement model equations possess the ability to reduce qualitative and quantitative variables to focus on their corresponding meaning. Therefore, one can use the constructs to condense information. For instance, strongly correlated processes can be abstracted, or specific environment variables can be aggregated.

Another important issue is that simulation data characteristics often violate the standard assumptions of statistical methods. This creates typical pitfalls in the analysis (Law 2014a), which can lead to misleading results. Ideally, residuals are NIID. However, NIID is not the default setting within a simulation. PLS-SEM is robust to NIID because a predictor specification only implies that construct scores are conditional expectations of the indicators. The specification can be expressed as

$$y_n = \alpha + \beta x_n + v_n \,. \tag{6}$$

Equation (6) is a generalized form for the two equation systems (see Section 2.1) for *n* observation points, which has the following implications: $E(v_n) = 0$, $\mathrm{cov}(x_n; v_n) = 0$ and $\mathrm{cov}(y_n; x_n) = \beta\,\mathrm{var}(x_n)$ (Lohmüller 1989). Eventually, this approach does not assume any statistical model and makes soft assumptions. In this way, it is a non-parametric approach, meaning that restrictive requirements do not apply here.

We will show that this condition also has useful implications for our metamodel's objective. For instance, a common pitfall of random simulation concerning NIID is the independence assumption of observations (Law 2014a). Considering equation (3), no correlation of $\mathrm{cor}(v_n, v_{n+1})$ is expected. Thus, PLS-SEM is also appropriate with the given dependencies of the observations. Another inherent characteristic is the heteroscedasticity of residuals. However, PLS-SEM does not expect identical residual distribution, besides $E(v_n) = E(v_{n+1}) = 0$. Furthermore, Lohmüller (1989) points out that neither the residual terms v_n and v_{n+1} nor the independent variables x_n and x_{n+1} need an identical distribution. For this reason, non-normal distributions do not necessarily lead to biased results (Reinartz et al. 2009).

To conclude, PLS-SEM is a complete non-parametric approach that possesses the capability to cover the emerging complex structures of models by also bearing many high-dimensional variables. Furthermore, because it respects the data characteristics, PLS-SEM leads to robust approximations (Henseler et al. 2009) that can be fairly suitable for understanding the behavior of simulation models.

2.3 SEM for Simulation Metamodeling

This subsection introduces the proposed metamodeling approach for simulation models by applying PLS-SEM. Once more, we describe the application of the two basic elements, the structural model and the measurement model for the simulation.

Concerning the structural model, we refer to the ontology of simulation models, according to which entities can be viewed as fundamental elements (Smith 2013). Thus, entities are represented as constructs in SEM. Additionally, mechanisms are modeled by groups of entities and their relationships (Hedström and Ylikoski 2010). A network of entities and structural relationships builds up the structural equation model and represents the underlying conceptual model of the simulation model by its semantic. The paths represent the structural assumptions, while the constructs embody the entities of simulation models. The semantic of SEM-links thus explicitly represents information on the simulation model and provides a conceptual model for investigating the behavior in an embedded scenario.

Having built up the elements and structures within the structural model, the measurement model is generated by assigning the simulation parameters as indicators to their corresponding constructs. Here, the best operationalization for the entities must be found, given the purpose of the analysis. This creates a number of challenges. For instance, an entity can sometimes be so complex that a few variables do not capture the content in all its facets. Another example could be a multivariate output of error measures. Single variables, which are sometimes the best expression of a simple entity, can also be used. In this case the construct becomes its measure (Diamantopoulos et al. 2012).

In this line, the following operationalization methods generally clarify the often missing link between entities and their corresponding measures. PLS-SEM offers two possible perspectives on this: reflective and formative constructs (see Subsection 2.1). These can be utilized for abstraction or aggregation for a set of variables, which may be important for a model's representation (Smith 2013). Therefore, it is possible to track the development from the conceptual to the mathematical simulation model by adding the measure to its corresponding construct.

Given the elements, structure, and measurements, the model can then be estimated. Recorded files containing simulated data from simulation experiments serve as data for the algorithm. Ideally, these simulated data are generated by applying a systematic experimental design (Barton 2013; Kleijnen and Sargent 2000; Law 2014b; Lorscheid et al. 2012).

The final step is to estimate and evaluate the metamodel. Given the parameter estimates from the PLS algorithm, the structural equation metamodel may be assessed and examined. The goal of the assessment is to evaluate the predictive capabilities and fit (Kleijnen and Sargent 2000) of the structures that are derived from the a priori information on the simulation model in combination with the measures calculated by the simulation data output. PLS-SEM possesses different established criteria for such an assessment. The validity and reliability of the constructs are evaluated by variance exploration. Reliability criteria such as Dillon-Goldstein's rho and Cronbach's alpha are typically implemented for evaluating the constructs (Nunnally and Bernstein 1994). Average variance extracted (AVE), which represents the proportion of the explained variance among the indicators, and cross-loadings assess aspects of validity. Variance inflation factors (VIFs) are also useful for detecting collinearity and avoiding violations. The structures can be evaluated by using complex resampling routines. Further, significances derived from several bootstrapping procedures, effect sizes (f^2), and the coefficients of determination (R^2) can be helpful for examination (see also Kleijnen and Deflandre 2006). Moreover, the Stone-Geisser resampling procedure is a cross-validation for drawing conclusions about prediction accuracy (Stone 1974).

Finally, the results of this metamodeling approach are communicated. PLS-SEM provides insights into existing structures and effects. By using the SEM semantic, we can present a conceptual model with specific simulation model information to reveal the structural assumptions and predictive causal relationships. Furthermore, the statistical analysis involves the simulation output from designed experiments to conduct a parameter estimation of the underlying relationships. Moreover, several possibilities even exist for the assessment of the metamodel's predictive power and fit. Finally, we present integrated information from the conceptual and data worlds within a structural equation metamodel that reveals the behavior of parts of the simulation model and offers convenient communication.

3 APPLICATION

This section presents and illustrates the proposed method by means of a simplified application. We use the learning agents for mechanism design analysis (LAMDA) simulation study by Lorscheid (2014). The model simulates a budgeting process that includes several steps, starting with reporting productivity values by department to achieve resources that are utilized to generate a certain department profit. The sum of the department profits results in a company profit. This process is embedded in an environment that affects the department in question. The research question in this sample model is about the drivers of a company's success, focusing on the effects of (un)truthful reporting.

First, we briefly introduce the variables. Since this model only serves as an illustration, an extensive description of the variables cannot be given here. For a more detailed overview, see Lorscheid (2014). A number of variables might influence a company's success. Discount limits the compensation payment to the departments based on the profit achieved. The variable OtherDepartments (OtherDep) measures other agents' reporting behavior, and Resources indicates the number of resources available to a company. A department's productivity level is operationalized by means of the following variables: Minimum represents the minimum level of productivity, whereas Scale determines its range. Reduction regulates the reduction of the departments' productivity above certain capacity levels for resource units. Agent behavior is determined by ReportValue, the reported productivity, and Deviation represents the deviation of the report from true productivity. A company's success is reflected in at least two variables: the dependent variables DivsionProfit (measuring the profit of the departments) and CompanyProfit (representing the overall company profit).

3.1 First-Order Polynomial Metamodel

In the next step, we calculate a first-order regression metamodel and present it as a metamodel for a simple input/output simulation model. This will be used as a reference point to elaborate on the contributions of the PLS-SEM metamodel. We do not use the equation expression, focusing instead on the impacts of the variables on the response.

Table 1 shows the results of the OLS regression estimation. Here, the values of standardized regression coefficients to their dependent variable are listed, assuming a direct relationship. Standardized regression coefficients are a convenient measure of a variable's impact. Generally, the regression fulfills its capability as a metamodel for the simulation model. Overall predictive capability can be expressed by R^2_{adj} and this is high for the two given metamodels, which indicates predictive power.

Table 1: Standardized OLS regression coefficients according to their response variables; not statistically significant = '*'.

Variables	DepartmentProfit (R^2_{adj} = 0.904)	CompanyProfit (R^2_{adj} = 0.805)
Minimum	0.012	0.108
Reduction	-0.069	-0.082
Scale	0.008	0.134
ReportValue	1.001	0.569
Deviation	-0.708	-0.441
Discount	-*	-*
OtherDepartments	-0.098	0.292
Resources	0.131	0.177

Given these results, one can conclude that ReportValue and Deviation are the main drivers of the dependent variables, according to the research question, while the impact of the productivity variables carries little weight. Further, considering the results of their p-values, ReportValue and Deviation have the strongest effects, and all variables are significant except Discount. In the following, these results are used as a reference point for the PLS-SEM analysis.

3.2 Structural Equation Metamodeling

In this subsection, we develop a structural equation metamodel with PLS-SEM following the structure presented in Subsection 2.3 to prepare the subsequent analysis of the sample model. First, we consider the a priori information provided by the simulation model and the research question. The sample model's purpose is to explore the impacts of several variables, including (un)truthful reporting, on company success. Thus, Reporting (REP) is defined as an independent entity-based construct. Entity success (SUC) is a dependent construct. Additionally, the model consists of a department environment and its attributes that affect overall and individual conditions. Therefore, we define the constructs for a department environment (DE) and department settings (DS) to examine their impacts. Having specified the model's constructs, we arrange their sequence along the implemented budgeting process.

We develop the paths based on the underlying theory. This allows us to shed light on the reporting behavior in its embedded environment and to disentangle the different cause/effect relationships of the independent variables. As one can see from Figure 2, the core relationship is REP's effect on SUC, since this reflects the research question, namely how reporting behavior influences success. The constructs DE and DS are implemented to analyze their influences on the relationship between REP and SUC. REP is influenced by the constructs DE and DS, while REP also affects SUC. This makes REP an intervening construct, allowing us to explicate the causal sequence between the variables (i.e. more than one layer of effects) and showing the path dependency. This finding shows the potential of our approach to make the interdependencies between the model elements explicit, visible, and communicable.

The following mathematical operationalization of the constructs can be viewed as a transition from a conceptual to a mathematical model. The constructs are linked to the input, output, and intermediate variables of the simulation model, as their measures in the structural equation model. Hence, we establish a close link between the mathematical and conceptual models. To link these measures with the constructs, we first organize our variables and constructs. The construct DE is determined by three measures, namely Discount, OtherDep, and Resources, which specify the department environment. The construct DS is caused by the three variables of Minimum, Scale, and Reduction, which limit the potential productivity of the department. REP is defined by the variables ReportValue and Deviation, which measure the reporting behavior of the department. We use the formative measure to include all its independent variance. The sum of department profits equals overall company profit. Therefore, DepartmentProfit and CompanyProfit are correlated and can be embodied in a reflective and dependent construct (SUC). The abstraction of these two subprocesses may reduce the model's complexity (i.e. from two models to one). Here, a reflectively measured construct is preferable (by integrating multiple output measures into one construct), because it can reduce the number of metamodels (Barton and Meckesheimer 2006). By running the PLS algorithm by means of SmartPLS (Ringle et al. 2014), we approximate the estimates of the metamodel (see Figure 2).

Generally, the structural equation metamodel can be read as follows. The rectangles represent the variables that are linked with their constructs, which are symbolized as circles. Therefore, this relationship is either a weight or the loading. The relationships between the constructs are the paths with their corresponding parameter estimates.

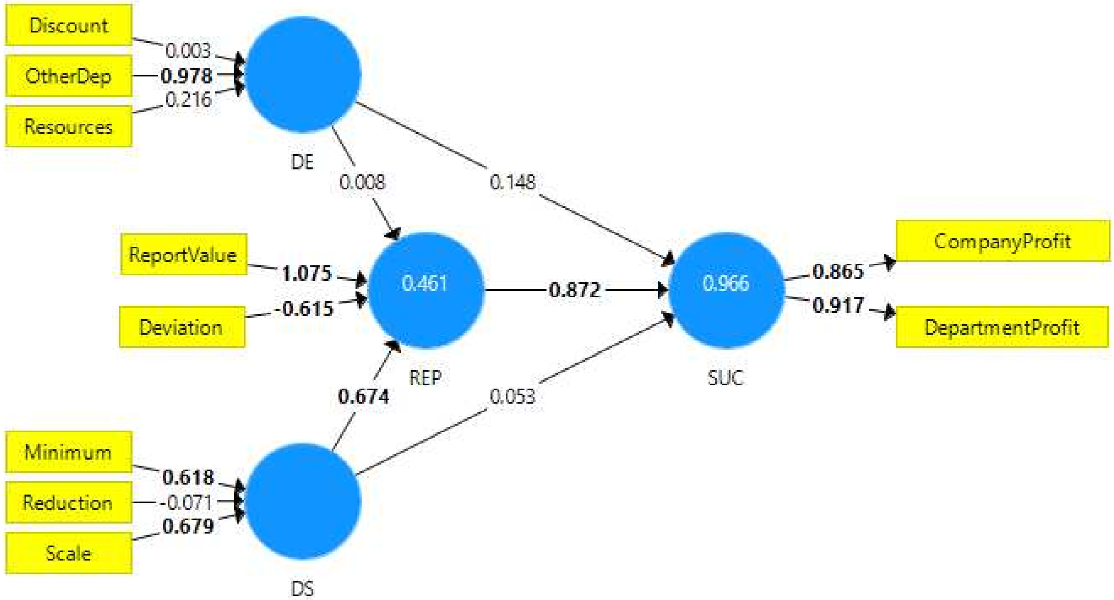

Figure 2: Structural Equation Metamodel.

Given the resulting metamodel, the structure and elements of the model and its central mechanisms are visible. This increases the transparency of the existing interdependencies and supports their communication. The overview of model elements and assigned input variables may thus serve as a basis for discussion.

3.3 Assessment of the Structural Equation Metamodel

The purpose of assessing the metamodel is to describe the structures by means of statistical measures and evaluate its predictive capability and the metamodel's fit to support and complement understanding (Kleijnen and Sargent 2000). As already described, an assessment of the measurement model can be performed by means of various non-parametric criteria. First, the reflective constructs are assessed; second, the formative constructs are analyzed; and finally, the model's structure is described.

The reflective constructs can be examined by their reliability and validity. First, the reliability of the construct SUC is investigated by means of Dillon-Goldstein's rho and Cronbach's alpha. The respective values are 0.855 and 0.855, which can be considered as reliable compared with the threshold of 0.8 (Nunnally and Bernstein 1994). Second, the AVE serves as a validity measure. The calculated AVE for the given construct is 0.794, which should at least equal 0.5. Other complementary measures are cross-loadings and the Fornell-Larcker criterion, which are not conspicuous in our metamodel (Fornell and Larcker 1981). In this way, we evaluate the proportion of indicator variance explained by its construct (Henseler et al. 2009). Overall, we can constitute that SUC is a well-defined target construct given the results from the analysis of the validity and reliability of the reflective constructs.

Because the formative constructs are causally determined by their indicators, the assessment has a different approach than that for reflective constructs. Primarily, these constructs should be tested chiefly with their conceptual meanings. The VIF measures then serve as the first assessment criterion to detect multicollinearity, because the measurement model is similar to multiple regression. The VIF values are below 5 for all formative constructs (DE, DS, REP) and model indicators. Therefore, strong collinearity can be excluded, suggesting the validity of the formative constructs (Henseler et al. 2009). As such, we conclude that the constructs are appropriate representations of our simulation entities.

Based on the assessed constructs, we can evaluate the structural model. Here, R^2_{adj} is one of the main assessment criteria for a metamodel's predictive capability. For the given model, the target construct SUC has an R^2_{adj} of 0.966. This finding indicates high predictive capability for the metamodel.

3.4 Comparison of the Metamodels

Both metamodeling techniques, OLS regression and PLS-SEM, show the crucial drivers of the simulation model. In particular for direct structures, low-order polynomial metamodels clarify the impacts of the variables, which may sufficiently represent simple models. However, if many variables and complex structures are involved, the balance can change.

First, the consideration of simulation model structures can lead to different conclusions concerning simulation behavior and results. In the given sample model, the relationships between DS, REP, and SUC are of special interest since they represent the research question. As mentioned, the low-order polynomial metamodel states that reporting is the most influential independent variable for company success. However, this conclusion does not match common sense. Company success will naturally be driven by productivity, which could lead to skepticism and misunderstanding. By means of our proposed method, one can see from the metamodel that there is in fact a mediation effect. Although DS has almost no direct effect on SUC, it strongly affects REP, which influences SUC. Hence, there is an indirect mediation effect of 0.588, which occurs by multiplying the direct paths from DS to REP (0.674) by SUC (0.872). This example shows how structural equation metamodeling may support the correct presentation of complex model behavior in a transparent and yet comprehensible way.

Second, the integrated information enables more insights into the behavior of simulation models. Structural equation metamodels enable an exploration of structures by strength and direction. In considering the relationship between DE and REP, the path has a coefficient value of 0.008. This value clearly indicates that there is only slight direct predictive causality, although this might be part of the simulation model structure. By contrast, REP has a high impact on success and has predictive relevance.

In this vein, the SEM semantic not only supports convenient communication but also checks the simulation model's plausibility.

Finally, PLS-SEM allows for the explicit analysis of complex structures and interdependencies. This is especially helpful when applied to more complex models. As one can see in Figure 2, we reduced several variables to constructs without changing the primary relationships and impacts. For example, we can consider ReportValue and Deviation in one construct or alone. In addition, using constructs enables interpretation and communication at the entity level by abstracting or aggregating the variables. Thus, complex structures such as entities with many variables can easily be reduced to constructs. Furthermore, by using constructs, it is possible to have one metamodel instead of several, thereby allowing a condensed presentation and communication of simulation behavior and results.

4 CONCLUSION AND FURTHER RESEARCH

We proposed an approach to the analysis of the behavior of simulation models and their subsequent communication by metamodels using PLS-SEM. Structural equation metamodels explicitly use a priori information from the conceptual model in combination with the simulation data output. Based on this, we estimated and evaluated the relationships and predictive capabilities of the metamodel. The link to theory considerably increases understanding compared with other metamodeling approaches. Overall, these structural equation metamodels may thus foster the development of the analytical expositions of simulation models while keeping them communicable and understandable.

This paper is a first step in this new direction, and more applications are required to test this approach's contribution. Indeed, PLS-SEM has been increasing its capabilities, such as consistent PLS (Dijkstra and Henseler 2015), which should also be explored by future research. Furthermore, SEM in general provides additional algorithms and goodness of fit measures that can be used to achieve other objectives in structural equation metamodeling, such as verification and validation.

REFERENCES

Barton, R. R. 2013. "Designing Simulation Experiments." In *Proceedings of the 2013 Winter Simulation Conference,* edited by S.-H. K. R. Pasupathy, A. Tolk, R. Hill, and M. E. Kuhl, 342-353. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., and M. Meckesheimer. 2006. "Metamodel-Based Simulation Optimization." In *Handbooks in Operation Research and Management Science,* edited by S. G. Henderson, and B. I. Nelson. Amsterdam, North-Holland: Elsevier B.V.

Byrne, B. M. 1998. *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. New York, New York: Psychology Press.

Diamantopoulos, A., M. Sarstedt, C. Fuchs, P. Wilczynski, and S. Kaiser. 2012. "Guidelines for Choosing between Multi-Item and Single-Item Scales for Construct Measurement: A Predictive Validity Perspective." *Journal of the Academy of Marketing Science* 40 (3):434-449.

Dijkstra, T. K., and J. Henseler. 2015. "Consistent and Asymptotically Normal PLS Estimators for Linear Structural Equation." *Computational Statistics & Data Analysis* 81 (1):10-23.

Falk, R. F., and N. B. Miller. 1992. *A Primer for Soft Modeling* Akron, Ohio: University of Akron Press.

Fornell, C., and F. L. Bookstein. 1982. "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory." *Journal of Marketing Research* 19 (4):440-452.

Fornell, C., and D. F. Larcker. 1981. "Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics." *Journal of Marketing Research* 18 (3):382-388.

Gerow, J. E., V. Grover, N. Roberts, and J. B. Thatcher. 2008. "The Diffusion of Second-Generation Statistical Techniques in Information Systems Research from 1990–2008." *Journal of Information Technology Theory and Application* 11 (4):5-28.

Hair, J. F., G. T. M. Hult, C. M. Ringle, and M. Sarstedt. 2014. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks, Los Angeles, California: Sage.

Hedström, P., and P. Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36 (1):49-67.

Henseler, J., C. M. Ringle, and R. R. Sinkovics. 2009. "The Use of Partial Least Squares Path Modeling in International Marketing." In *New Challenges to International Marketing (Advances in International Marketing, Volume 20)*, edited by R. R. Sinkovics, and P. N. Ghauri, 277-319. Bingley, UK: Emerald Group Publishing Limited.

Kleijnen, J. P. C. 2014. "Simulation-Optimization Via Kriging and Bootstrapping: A Survey." *Journal of Simulation* 8 (4):241-250.

Kleijnen, J. P. C., and D. Deflandre. 2006. "Validation of Regression Metamodels in Simulation: Bootstrap Approach." *European Journal of Operational Research* 170 (1):120-131.

Kleijnen, J. P. C., and R. G. Sargent. 2000. "A Methodology for Fitting and Validating Metamodels in Simulation." *European Journal of Operational Research* 120 (1):14-29.

Law, A. M. 2014a. *Simulation Modeling and Analysis*. Vol. 5th International Edition. New York, New York: McGraw-Hill, Inc.

Law, A. M. 2014b. "A Tutorial on Design of Experiments for Simulation Modeling." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 66-80. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lohmüller, J. 1989. *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg, Germany: Physica-Verlag.

Lorscheid, I. 2014. "LAMDA - Learning Agents for Mechanism-Design Analysis ", Institute of Management Control and Accounting, Hamburg University of Technology, Hamburg, Germany [Accessed September 26th, 2014].

Lorscheid, I., B.-O. Heine, and M. Meyer. 2012. "Opening the 'Black Box' of Simulations: Increased Transparency and Effective Communication through the Systematic Design of Experiments." *Computational and Mathematical Organization Theory* 18 (1):22-62.

Lowry, P. B. 2014. "Partial Least Squares (PLS) Structural Equation Modeling (SEM) for Building and Testing Behavioral Causal Theory: When to Choose It and How to Use It." *IEEE Transactions on Professional Communication* 57 (2):123-146.

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

Reinartz, W., M. Haenlein, and J. Henseler. 2009. "An Empirical Comparison of the Efficacy of Covariance-Based and Variance-Based SEM." *International Journal of Research in Marketing* 26 (4):332-344.

Ringle, C. M., S. Wende, and J. Becker. 2015. "SmartPLS 3." *Hamburg: SmartPLS: Retrieved from http://www.smartpls.com* [Accessed March 31st, 2015].

Sanchez, S. M., and H. Wan. 2012. "Work Smarter, Not Harder: A Tutorial on Designing and Conducting Simulation Experiments." In *Proceedings of the 2012 Winter Simulation Conference*, edited by J. H. C. Laroque, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1929-1943. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sargent, R. G. 1991. "Research Issues in Metamodeling." In *Proceedings of the 1991 Winter Simulation Conference*, edited by B. L. Nelson, W. D. Kelton, and G. M. Clark, 888-893. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sarstedt, M., C. M. Ringle, J. Henseler, and J. F. Hair. 2014. "On the Emancipation of PLS-SEM: A Commentary on Rigdon (2012)." *Long Range Planning* 47 (3):154 -160.

Smith, R. 2013. "On the Value of a Taxonomy in Modeling." In *Ontology, Epistemology, and Teleology for Modeling and Simulation - Philosophical Foundations for Intelligent M&S Applications,* edited by A. Tolk, 241-254. Heidelberg, Germany: Springer.

Stone, M. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2):111-147.

Sturrock, D. T. 2014. "Tutorial: Tips for Successful Practice of Simulation." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 90-97. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Tolk, A. 2013. "Truth, Trust, and Turing - Implications for Modeling and Simulation." In *Ontology, Epistemology, and Teleology for Modeling and Simulation - Philosophical Foundations for Intelligent M&S Applications*, edited by A. Tolk, 1-26. Heidelberg, Germany: Springer.

Tolk, A., B. L. Heath, M. Ihrig, J. J. Padilla, E. H. Page, E. D. Suarez, C. Szabo, P. Weirich, and L. Yilmaz. 2013. "Epistemology of Modeling and Simulation." In *Proceedings of the 2013 Winter Simulation Conference*, edited by S.-H. K. R. Pasupathy, A. Tolk, R. Hill, and M. E. Kuhl, 1152-1166. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wold, H. 1982. "Soft Modeling: The Basic Design and Some Extensions." In *Systems under Indirect Observation, Part II*, edited by K. G. Jöreskog, and H. Wold, 1-54. Amsterdam, North-Holland: Elsevier Science Ltd.

Yilmaz, L., S. J. E. Taylor, R. Fujimoto, and F. Darema. 2014. "Panel: The Future of Research in Modeling & Simulation." In *Proceedings of the 2014 Winter Simulation Conference* edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 2797-2811. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

KAI G. MERTENS is an associate researcher at the Institute of Management Control and Accounting at the Hamburg University of Technology, under the mentorship of Prof. Dr. habil. Matthias Meyer. He has a Master’s degree in industrial engineering and has worked as an electronics engineer. His email address is kai.mertens@tuhh.de.

IRIS LORSCHEID is a postdoctoral researcher and lecturer at the Institute of Management Control and Accounting at the Hamburg University of Technology. Her research interests include agent-based modeling and simulation model analysis. She received her diploma in computer science from the University of Koblenz and her PhD (Doctorate in Natural Sciences) from the Hamburg University of Technology. Her email address is iris.lorscheid@tuhh.de.

MATTHIAS MEYER is a professor of management control and accounting and Head of the Institute of Management Control and Accounting at the Hamburg University of Technology. His research interests include simulation data analysis, simulation modeling, the design of cost accounting systems, and modeling of accounting routines. He has a habilitation degree from the WHU – Otto Beisheim School of Management and he received a PhD from Ludwig-Maximilians-University in Munich in 2003 for his dissertation on Principal Agent Theory and Methodology. He received a diploma in business administration from the Catholic University of Eichstätt-Ingolstadt and a Master’s in philosophy of the social sciences from the London School of Economics. His email address is matthias.meyer@tuhh.de.