Chapter 3

Predictions from Partial Least Squares Models

Nicholas P. Danks and Soumya Ray

1. Introduction

This chapter seeks to introduce partial least squares (PLS) practitioners to the generation and evaluation of predictions from their path models, both as a means of validating the practical usefulness of their models and for forecasting future outcomes.

Studies in tourism and hospitality currently offer strong research utility by explaining how personal, regional, or commercial factors generally relate to positive outcomes for the industry. However, the practical utility of our studies is limited to general policy-making suggestions based on the significance of antecedent factors. Could we use our models to predict, say, a new subject's potential behavior in the future, or forecast how a particular region or commercial interest will fare in the coming year given new data? In addition, before we even have access to data on new subjects, can we use our existing samples to validate that our current models have the potential to make accurate predictions using new data? Being able to address these types of prediction problems would greatly magnify the practical utility of our models and better serve the immediate and quantifiable needs of governments, organizations, businesses, services, and persons involved in the tourism industry. Current approaches and metrics in PLS path modeling (PM) largely cannot answer these prediction-oriented questions about specific cases or scenarios, and are largely relegated to making highly generalized prognostications based on path significances.

We believe that predictive PLS is one of the emerging and promising directions in PLS-PM. Although predictive techniques for PLS are still at a nascent stage of development, we hope this chapter brings you up to speed on the latest developments, informs you of predictive practices you can employ today, and gives you a foundation for following future developments in this exciting new direction for PLS-PM.

Applying Partial Least Squares in Tourism and Hospitality Research, 35–52

doi:10.1108/978-1-78756-699-620181003

We start by examining some of the general terminologies and techniques pertaining to predictive modeling, as illustrated in Fig. 1. *Prediction*, especially in the context of regression-based techniques, is the use of a dataset and an empirical model to predict unknown outcomes. We typically train a predictive model from existing predictor and outcome variables, such that it can use new predictor variables to predict new outcomes.

Training a predictive model involves estimating model parameters with the intention to use them to predict outcome variable data from predictor data. *Training data* refer to the data used in the parameter estimation needed for model training and is often called in-sample (IS) data. A *trained model* is a model, which has used the training data to estimate parameters and can be used for prediction. *Holdout data* are predictor data that were not used in model training and are intended for use in generating predictions on outcomes variables – these are often called out-of-sample (OOS) data. *Predicting*, in this chapter, is applying a trained model on holdout predictor variables to generate predicted values for outcome variables of interest. In contrast to prediction, *fitting* refers to using a trained model to regenerate outcomes for the training data itself.

Purely predictive models seek to extract all predictive information from training data regardless of theoretical validity of predictors or relationships. This purely predictive approach typically ignores a priori theory and often, but not always, seeks to make case-wise predictions rather than average-case predictions. Examples of such purely predictive models include random forests and neural networks. Such predictions can have practical implications when the model is being used to prepare for future actions.

We can also be interested in making predictions where actual outcomes are already known, when we have both the predictor and outcome variable holdout data.

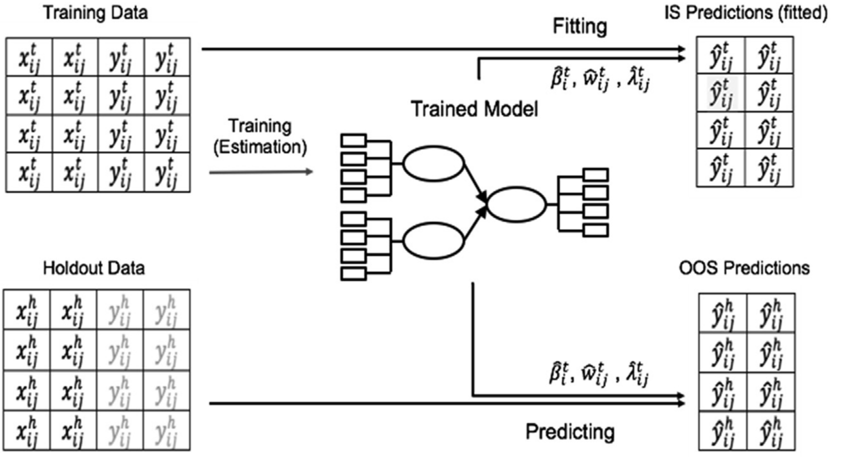

Fig. 1: An Overview of the Prediction Process. x_{ij} Are Predictor Items, y_{ij} Are Actual Outcome Items, $\hat{y}_{ij}$ Are Predicted Values of Outcome Items, and $\hat{w}^t_{ij}$, $\hat{\beta}^t_{ij}$, and $\hat{\lambda}^t_{ij}$ Are the Measurement and Structural Estimates of the PLS-PM.

These predictions can be used to validate the predictive qualities of the model and demonstrate how well the model might predict the unseen data. If predicted outcome values are similar to known outcome values for a given set of predictors, we can conclude that the model predicts the phenomenon well. This form of *predictive validity* has practical value because it lets us determine whether a model can offer specific prescriptions for various managerial and practical situations.

PLS prediction serves two essential purposes: generating theory-driven predictions and gauging the predictive validity of theoretical models. In generating predictions for unknown outcomes, PLS prediction differs greatly from the purely predictive methods discussed earlier. PLS prediction utilizes a theory-driven, expert-built model to predict outcomes. Thus, PLS prediction does not optimize for the best predictive power – the model is informed by theory and hypothesis and then prediction is performed after the fact. Furthermore, PLS path models model average-case behavior; thus, predictions from PLS path models generate average-case predictions for a given set of predictors.

Shmueli and Koppius (2011) describe various roles for predictive analytics in scientific research and argue for a joint explanatory and predictive approach. Although the approach of explanation first and prediction second limits the predictive power of the model, it permits us to determine the predictive validity of a theory. This approach conforms to the joint explanatory and predictive approach recommended by Shmueli and Koppius (2011).

The PLS-PM literature has called for increased emphasis on the evaluation of predictive performance in assessing PLS models (Dijkstra, 2010; Hair, Ringle, & Sarstedt, 2011) and we expect that journal editors and reviewers will increasingly expect the evaluation of predictive validity in PLS new submissions.

The literature on predictive PLS has also discussed what constitutes a good predictive metric and how to evaluate predictive validity (e.g., Chin, 2010; Evermann & Tate, 2012, 2014, 2016). We will follow the very recent recommendations of Shmueli, Ray, Estrada, and Chatla (2016), who outline an algorithm for generating PLS predictions and suggest the use of OOS aggregate error metrics such as root mean square error (RMSE) and mean absolute error (MAE) for the endogenous composite indicators.

In this chapter, we aim to provide a theoretical foundation of the latest developments in PLS prediction, provide guidelines on the application and evaluation of predictions, and provide an applied empirical example employing the techniques described in the chapter. It is our hope that this chapter can be of service to members of the PLS community wishing to apply prediction in their research.

2. How to Predict from PLS Models

Unlike covariance-based SEM, PLS-PM has interesting properties that allow us to generate predictions. PLS primarily estimates composite measurement models rather than common factors, unless we make special adjustments to do so (Henseler et al., 2014; Rigdon, 2012; Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016). Composite constructs are weighted sums of their indicators, using weighting modes such as mode A (correlation weights), mode B (regression weights), or weighted sum (unit weights). Note that both weights and loadings are always estimated for each indicator of a composite and incorporate the measurement mode specified. Thus, composites have a determinate score, which can be estimated from the standardized indicator scores and indicator weights for any specified measurement mode.

Currently proposed schemes to generate predictions from PLS are largely limited to composites, because composite scores and the causal structure between composites allow us straightforward ways to predict outcome scores and indicators regardless of the measurement mode. Although future prediction schemes might incorporate common factors, our discussion here will advocate the use of composites for ease of implementation and interpretation of predictions.

2.1. Generating Predictions from PLS-PM

We briefly summarize and describe the PLSpredict algorithm as proposed by Shmueli et al. (2016), which describes a straightforward method for prediction that others have likely also implemented (e.g., Evermann & Tate, 2016). We refer to Fig. 2 that demonstrates a simple PLS model for the discussion of the predictive algorithm.

1. *Randomly partition the data.* If only one dataset is available, randomly partition the data into a training subset and holdout subset. If a training set and

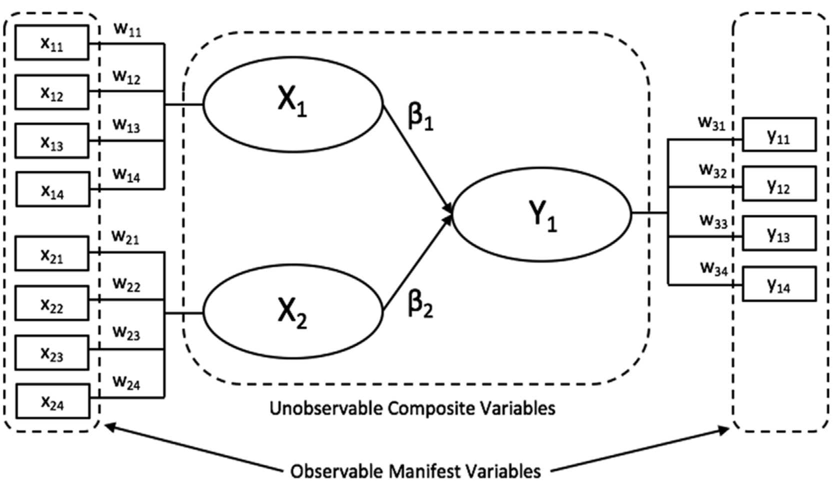

Fig. 2: A Simple PLS-PM Composite Model. x_{ij}: Observable Measures of Exogenous Composites, w_{ij}: Measurement Weights, X_i: Exogenous Composites, β_i: Path Coefficients, Y_i: Endogenous Composites, and y_{ij}: Observable Measures of Endogenous Composites.

holdout set are already at hand - if for instance, data was collected twice during the empirical study - this step is not necessary.
For each manifest variable, we will thus have x_{ij}^{t} and x_{ij}^{h} with the superscript *t* referring to training subset and *h* to holdout subset, and indices *i* to the composite index and *j* to the indicator index for that composite. x_{11}^{t} is thus the first training data subset indicator of the composite X_1; and x_{11}^{h} its holdout set counterpart.

2. *Estimate the training model.* Using data only from the *training* set, estimate the PLS model.
Retain all training model weights $\left(w_{ij}^{t}\right)$, loadings $\left(\lambda_{ij}^{t}\right)$, path coefficients $\left(\beta_{i}^{t}\right)$, and the standard deviation $\left(sd_{ij}^{t}\right)$ and mean $\left(m_{ij}^{t}\right)$ for each of the indicator scores in the training data, respectively.
3. *Predict the antecedent composite scores from antecedent indicators.* Using data only from the holdout set and the estimated parameters from the training model.
Standardize the holdout antecedent indicator scores by deducting the mean $\left(m_{ij}^{t}\right)$ and dividing by the standard deviation $\left(sd_{ij}^{t}\right)$ for each corresponding indicator score from the training set.

$$x_{ij}^{h*} = \left(x_{ij}^{h} - m_{ij}^{t}\right) / \, sd_{ij}^{t}$$

Predict the antecedent composite score by multiplying standardized antecedent indicator scores by their respective training weights and summing across indicators.

$$\hat{X}_{i}^{h} = \sum\nolimits_{j=1}^{4}\left(x_{ij}^{h*} \times \omega_{ij}^{t}\right), i = 1,2$$

4. *Predict the endogenous composite scores.* Multiply the predicted antecedent composite scores with their respective path coefficients and summing for each observation.

$$\hat{Y}_{i}^{h} = \sum\nolimits_{i=1}^{2}\left(\hat{X}_{i}^{h} \times \beta_{i}^{t}\right)$$

5. *Predict the OOS indicator scores of the endogenous composites.* Multiply the predicted antecedent composite score by each respective loading (training) to obtain each respective composite score.

$$\hat{y}_{ij}^{h*} = \hat{Y}_{i} \times \lambda_{ij}^{t}$$

De-standardize the indicator scores by multiplying by the training SD and adding back the training mean for each corresponding indicator from the training set.

$$\hat{y}_{ij}^{h} = \left(\hat{y}_{ij}^{h*} \times \, sd_{ij}^{t}\right) + m_{ij}^{t}$$

6. *Predict the IS indicator scores of the endogenous composites***.** Repeat Steps 3 and 4 using training indicator data $\left(x_{ij}^{t}\right)$ to generate predicted indicator scores of the outcome composite for training data $\left(\hat{y}_{ij}^{t}\right)$.

7. *Calculate predictive metrics***.** Metrics are calculated using the predictive error, which is generated by deducting the predicted value from the known value for both the IS predicted indicator scores $\left(y_{ij}^{t} - \hat{y}_{ij}^{t}\right)$ and OOS predicted indicator scores $\left(y_{ij}^{h} - \hat{y}_{ij}^{h}\right)$.

2.2. Earliest Versus Direct Antecedents Approaches

Shmueli et al. (2016) acknowledge the shortcomings of the PLSpredict algorithm when applied to a model including one or more mediating constructs. Mediators pose a special challenge in the predictive context, in that their composite scores can be both predicted by antecedent composites or by the composite's own indicators and training weights; however, only one of these composite scores can be used in the predictive algorithm. Shmueli et al. (2016) propose either treating the mediator as a purely intervening variable and use the earlier antecedents to predict composite score of mediators; or treating mediators as purely exogenous constructs (thus, dropping earlier antecedent constructs) and using their indicator scores and training weights to generate its composite score and then directly predict outcome constructs. Refer to Fig. 3.

We refer to the approach that treats the mediators as intervening variables, and predicts the composite score from only earliest exogenous antecedents, as the earliest antecedents (EA) approach. In contrast, we refer to the approach that treats the mediator as exogenous variable only, and uses their indicators and weights to generate composite scores that are then used in predicting outcomes, as the direct antecedents (DA) approach. Both techniques have shortcomings in that they necessarily ignore some part of the measurement model – the indicators and weights of the mediator in the case of the EA approach, or the indicators and weights of earlier antecedents in the case of the DA approach.

There are valid considerations and applications for both the EA and the DA approaches. The PLS estimation algorithm treats mediators as purely exogenous variables when estimating its paths to their outcome constructs. Thus, the DA approach is more true to the prediction of outcomes and should be more accurate than EA prediction largely.

However, if the data are longitudinal rather than cross-sectional in nature, we might currently only have data on exogenous antecedents and not for mediators when doing prediction. In such cases, we would need to predict both mediating and endogenous composites and the EA approach would be more practical and suitable for such predictions and forecasts. We recommend that practitioners be given the choice of predictive scheme to be applied. To the best of our knowledge, SmartPLS V3.2.7 (Ringle, Wende, & Becker, 2015) has implemented the EA approach.

3. Evaluating Predictions from PLS Models

For practical and useful evaluation of predictions, we limit our discussion to OOS, operative prediction, which best fits the conceptual vision of prediction of the large data analytics community (Shmueli et al., 2016).

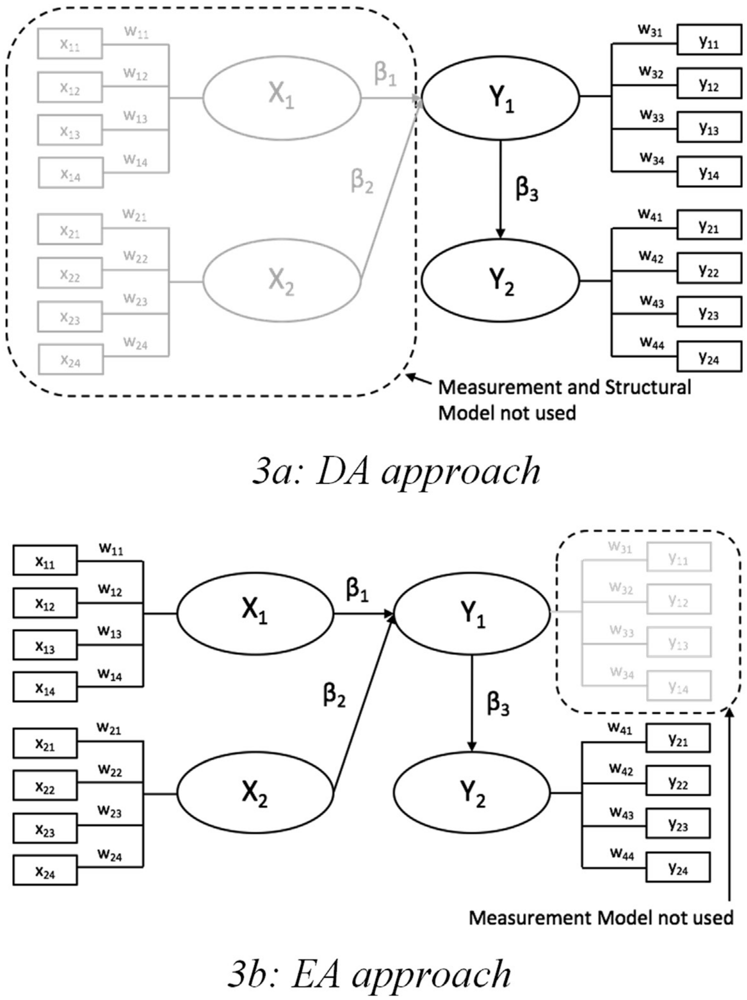

Fig. 3: Information Disregarded in Direct Versus Earliest Antecedents Approaches. Grayed-Out Areas of the PLS path model Demonstrate Data and Parameter Estimates Not Used in the Prediction of Y_1 or Y_2.

3.1. In-sample Versus Out-of-sample

Using the PLS prediction methods outlined earlier, domain experts can evaluate the practical predictive utility of their models, backed by theoretical rigor. Evaluating the predictive qualities of models informs researchers of overfit to the sample data at hand. Models with good predictive qualities can be used to predict outcomes for new subjects or cases. Thus, gauging the extent of overfit can help practitioners, and their readers, judge the practical utility of their models for prediction.

While evaluating the predictive overfit of PLS models, it is important to distinguish between the predictive power of IS (fitted) cases versus new, OOS (predicted) cases. When predictions are made on the dataset upon which the model was estimated, the resultant predictions are IS predictions and can be seen as an evaluation of the fit (in this case, the predictions are called the fitted values). This estimated training model will capture both the signal in the data and some of the noise; thus, "predictive" performance could be overstated or biased in that it is not seeing new signal or noise. When predictions are made on a new unseen (or holdout) dataset, the resultant predictions are OOS predictions and can be seen as an evaluation of the ability of the estimated model to predict the signal in the new dataset.

If we have both IS and OOS predictive metrics, we can evaluate whether predictive performance increases or decreases comparing training outcome predictions to the holdout outcome predictions. Refer to Table 1 for a comparison.

If there is little decrease in predictive power (which we expect) – or little increase in predictive power (which we do not expect) – then the model is not overfit. If there is a dramatic decrease in the predictive power of holdout outcomes, the model is not suitably predicting the signal in the holdout dataset. This might be due to inherent differences in the signal in the data or the model being overly sensitive to the noise in the training dataset. If there is a dramatic improvement (underfit), the signal in both the training and holdout are captured but the training set might be noisier or contain a weaker signal.

3.2. Cross Validation

Cross validation is reusing the total dataset by partitioning the available sample into subsets of training and test. Common cross-validation techniques include *k*-fold and leave-one-out. Cross validation is good when samples are small and a simple partition would leave little data for both training the model and evaluating the predictions in the holdout set. Cross validation can also give a better idea of practical OOS performance when data cannot be spared for a test set.

k-Fold cross validation is easy to implement, allows for each observation to be included in the test set one time, and is not computationally intense. We therefore recommend the use of the *k*-fold technique. However, there are many alternatives and we recommend Hastie, Tibshirani, and Friedman (2009) for further details.

3.3. Predictive Metrics

It is important to identify which predictive metrics are best suited to evaluate the predictive power of the model. We consider the most popular generic metrics being RMSE, MAE, mean bias, and R^2.

RMSE is a widely used predictive metric (Hastie et al., 2009) and highly interpretable. RMSE is in the original scale of the data and can be seen as the standard deviation of the predictive error. Thus, a smaller RMSE value will indicate a

Table 1: Comparison of IS Versus OOS Predictive Metrics for the Evaluation of Predictive Validity.

Condition	Test RMSE Low	Test RMSE High
Train RMSE Low	Good predictive validity	Overfit
Train RMSE High	Underfit	Worrisome predictive validity

tighter fit of the predictions to the true data. RMSE squares the prediction error and thus a larger penalty is paid for extreme errors.

MAE is a measure of the average absolute predictive error. MAE is in the original scale of the data and tells us on average how different our predictions are from the true known values. Unlike RMSE, there is no additional penalty for extreme errors.

Both RMSE and MAE treat over- and under-prediction identically and give no clear indication of bias. Therefore, we recommend that one visualizes the error in a residuals plot and calculates the mean error for indications of systematic bias in the predictions and predictive error.

R^2 is a much-reported metric in the literature; however, it suffers from some drawbacks within the predictive context. R^2 is the IS squared correlation between fitted and true composite scores. Thus, R^2 is evaluating the IS predictive performance, but does not provide information on how well the model performs on unseen holdout data. IS metrics are of value in evaluating predictive power, but should not be conflated with true OOS predictive metrics (Shmueli et al., 2016).

3.4. Benchmarks

In evaluating the performance of any model, predictive methodology suggests the use of a naive benchmark. In line with the concept of scientific parsimony, the model proposed should provide increased predictive power for added complexity. A naive benchmark is essentially a target for the minimum predictive power we wish to exceed. Often, pure prediction practitioners will use the training set mean value as a naive prediction or the predictions generated by a simple alternate algorithm to serve as the naive benchmark.

A naive predictive benchmark of the training mean might be viewed as insufficiently rigorous. Therefore, we will discuss the use of a linear model as a comparative predictive technique (Evermann & Tate, 2016; Shmueli et al., 2016) to serve as benchmark comparison. Evermann and Tate (2016) and McDonald (1996) note that due to the constraints imposed by the measurement and structural models of PM, a simple linear regression (LM) of the endogenous indicator variables on the exogenous indicator variables should provide a better prediction than PLS-PM. However, Evermann and Tate (2016) find in their simulation study that this is not supported – "all PLS-PM methods perform slightly better than linear multiple regression." The LM thus serves as a good benchmark with which to compare predictive power – the PLS path model's predictive power should be at least equal to that of the LM, with larger improvements demonstrating increasing predictive power.

4. Evaluating Predictive Performance

The PLSpredict algorithm provides a technique for generating predictions from the estimated model and predictive metrics for the evaluation of these predictions. This gives us a specific idea of how well we can predict the indicator. In the case of Likert-type survey data – often used in tourism research – it allows us to predict a specific case response for the Likert survey question. This is of great value when considering the practical applications of the research and for making managerial recommendations.

It is important to note that the context and domain play an important role in the evaluation of the results. The question of what an acceptable level of predictive error for the theory/context is is heavily dependent on the application and domain. For example, lower predictive power might be more acceptable in social sciences than in medicine, where lives might be at stake. It is also important to note that there might be a difference in cost of over- and under-prediction. Over-predicting staff turnover might have a lower cost than under-predicting while the opposite might affect customer satisfaction.

The bias in the predictive error will provide evidence of whether the model is systematically over- or under-predicting. It is useful to visualize the predictive error in a residuals plot. This plot should follow a standard normal distribution – skewed error provides the evidence of systematic bias. If the model is systematically biased on new data, it will tell us that the estimated model does not have sufficient predictive power for the prediction of these cases.

Predictive metrics are useful in identifying specific cases of interest. RMSE gives us the standard deviation of the predictive error. A distribution of predictive errors can be plotted and errors outside a chosen prediction interval can be identified. These cases can be seen as not being predicted well by the model and with access to the actual values for these outlier cases, we can begin to investigate why they are suffering from poor prediction – this has the potential to inform modeling or sample design.

If a mediator is of central importance to the empirical theory, we can leverage the use of EA and DA approaches to the predictive algorithm to lend an additional insight into the efficacy of the mediator. EA predictions can be compared to the DA predictions – this constitutes a comparison of how well the EA composites predict the outcome compared to the mediating composite and provides additional insight as to the loss or gain in predictive power attributable to the mediator.

5. Using PLSpredict (An Empirical Example)

We now present a demonstration of the predictive techniques described in this chapter using an empirical model from the hospitality and tourism literature – *Effects of*

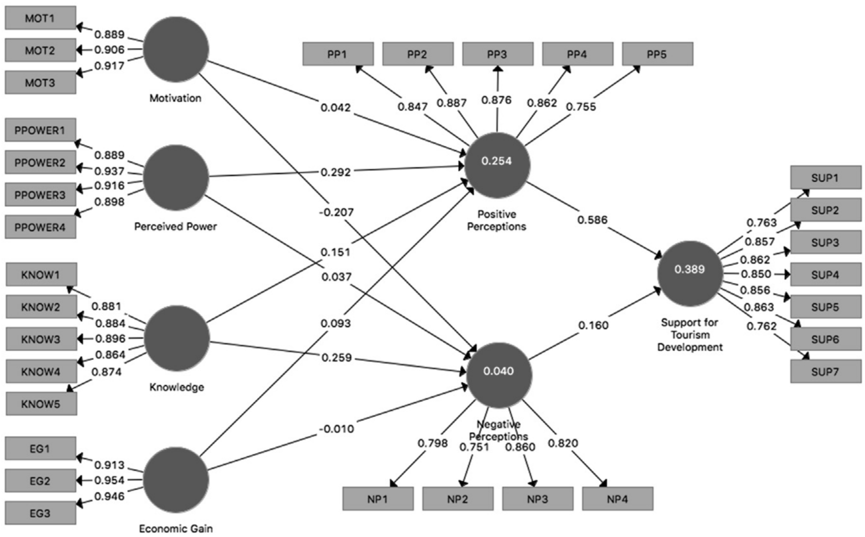

Fig. 4: The PLS Model Adapted from Rasoolimanesh et al. (2017).

Motivation, Knowledge and Perceived Power on Residents' Perceptions: Application of Weber's Theory in World Heritage Site Destinations (Rasoolimanesh, Jaafar, & Barghi, 2017).

The data comprise survey data collected from urban residents of the George Town World Heritage Site, Malaysia. The structural model consists of seven construct variables. These have four antecedent constructs: motivation of residents to become involved with and contribute to tourism development (Motivation), perceived power to make decisions and control the process of tourism development (Perceived Power), knowledge and awareness regarding tourism development and its impacts on their community (Knowledge), economic gain and benefits from tourism (Economic Gain). Two fully mediating constructs: positive perceptions (Positive Perceptions, PP) and negative perceptions (Negative Perceptions, NP) toward the impacts of tourism development. Last, one outcome construct: support for tourism development in their community (Support for Tourism Development). All constructs are estimated using measurement mode A and have between three and seven items each. The authors envision the constructs as reflective factors; however, for the purposes of prediction, we will treat them as composites with determinable scores.

The study was originally run in SmartPLS V3.2.3 but we used the latest version available at the time of our writing, SmartPLS V3.2.7, to employ the analysis. The model estimates are identical across software versions (Fig. 4).

5.1. Predictive Value of Model

As discussed in Section 3.4, we will first consider the model from the standpoint of predictive parsimony. This will assess whether the additional complexity of the measurement and structural model result in an adequate increase in the predictive power. To achieve this, we perform, 10-fold cross validation in SmartPLS V3.2.7 to generate OOS RMSE and MAE metrics for both the linear model, and the PLS path model for endogenous composite indicators (Table 2).

The indicators of the NP construct show very little or no improvement in predictive power of the PLS model over the LM benchmark. This is not surprising given the low R^2 of the NP construct (R^2 = 0.04) – if IS average-case predictive power is low (R^2), one would expect OOS predictive power to be similarly low. The remaining endogenous constructs PP (R^2 = 0.254) and Support for Tourism Development (R^2 = 0.389) demonstrate better improvement of predictive power over the LM benchmark, especially for MAE. This improvement in predictive power of the PLS model above the LM demonstrates that the PLS model has sufficient predictive power to support the use of the predictive PLS algorithm over more parsimonious techniques and supports the use of the predictive model.

It is also important to evaluate the indicator-level RMSE in the *scale* of the indicator. That is, by construction the indicators of Support for Tourism Development (SUP1–SUP7) are measured on a 5-point likert scale and have an RMSE of 0.706–0.779. We can thus say that on average 68% (1 SD) of prediction errors will fall within 1.5 points of the 5 points scale. If the true value is 3, 68% of predictions fall between 2.27 and 3.73, and 95% of predictions fall between 1.54 and 4.46. This range represents nearly the full range of the indicators' measurement scale and one needs to consider if this is acceptable given the context.

5.2. Predictive Value of Proposed Versus Alternative Models

As discussed earlier, if mediators are an important contribution of the article, one needs to also consider the gain or loss in predictive power of the EA approach (which considers the predictive power of earlier antecedents) versus the DA approach (which considers the predictive power of the mediators directly).

To achieve this in SmartPLS V3.2.7, which does not provide the choice of EA or DA as predictive technique, we simply performed PLS prediction first with the full model, including the EA (Motivation, Perceived Power, Knowledge, and Economic Gain); and second with the EA removed, such that only the mediators were predicting Support for Tourism Development. It needs to be noted that at this point, such a comparison does not preclude the need for formal tests of mediation, but serves only to gauge the gain or loss in predictive power due to the mediator. We expect that a similar technique will present itself for ease-of-use in leading PLS software in the coming years.

In Table 3, we present the predictive power of EA versus DA approaches. We can see the improved predictive power of the outcome indicators using the mediator versus earlier antecedents by considering the percentage improvement in both RMSE and MAE for the DA approach. All values are positive indicating a lower RMSE and MAE for the predictions generated by the mediators and demonstrate a large improvement in predictive power of 7.14–17.95%. We thus conclude that the mediators are better predictors of the outcome indicators. This demonstrates that the mediators are of predictive value in the model – they predict the outcome better than the antecedents do.

Table 2: Predictive Performance of the PLS Model Versus Benchmark LM.

		PLS Predict		LM Predict		LM-PLS		(LM-PLS)/PLS (%)	
Composite	**Indicator**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
Negative Perceptions	NP1	1.323	1.147	1.337	1.141	0.014	−0.006	1.05	−0.53
	NP2	1.222	0.984	1.230	0.990	0.008	0.006	0.69	0.57
	NP3	1.354	1.182	1.374	1.176	0.019	−0.006	1.44	−0.50
	NP4	1.300	1.119	1.295	1.083	−0.005	−0.036	−0.41	−3.21
Positive Perceptions	PP1	0.726	0.562	0.741	0.580	0.015	0.018	2.09	3.27
	PP2	0.691	0.529	0.710	0.545	0.019	0.017	2.81	3.14
	PP3	0.721	0.570	0.738	0.582	0.016	0.012	2.28	2.12
	PP4	0.739	0.572	0.750	0.583	0.011	0.011	1.52	1.89
	PP5	0.671	0.508	0.695	0.519	0.024	0.011	3.54	2.23
Support for Tourism Development	SUP1	0.779	0.535	0.802	0.576	0.023	0.041	2.92	7.66
	SUP2	0.751	0.533	0.743	0.554	−0.008	0.021	−1.01	3.90
	SUP3	0.719	0.507	0.722	0.532	0.003	0.024	0.41	4.78
	SUP4	0.711	0.512	0.720	0.548	0.008	0.036	1.18	7.09
	SUP5	0.714	0.532	0.720	0.559	0.006	0.027	0.82	5.09
	SUP6	0.706	0.519	0.713	0.542	0.007	0.023	1.05	4.37
	SUP7	0.747	0.538	0.772	0.567	0.025	0.029	3.35	5.44

Notes. Gray-shaded values demonstrate indicators for which there is no improvement in predictive power of the PLS model over the LM benchmark.

Table 3: Predictive Power of Mediator Versus Earlier Antecedents.

		EA Approach		DA Approach		(EA-DA)		(EA-DA) / DA (%)	
Composite	**Indicator**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
Support for Tourism Development	SUP1	0.779	0.535	0.726	0.487	0.053	0.047	7.29	9.74
	SUP2	0.751	0.533	0.663	0.452	0.088	0.081	13.20	17.95
	SUP3	0.719	0.507	0.653	0.442	0.066	0.066	10.14	14.87
	SUP4	0.711	0.512	0.642	0.447	0.069	0.065	10.79	14.42
	SUP5	0.714	0.532	0.666	0.470	0.048	0.062	7.14	13.10
	SUP6	0.706	0.519	0.653	0 463	0.053	0.056	8.12	12.15
	SUP7	0.747	0.538	0.689	0.482	0.057	0.056	8.33	11.56

Notes. EA-DA tells us the predictive penalty of not having a mediator in the model; we can also interpret this in relative sense as a percentage penalty in predictive power.

5.3. Evaluating the Prediction Residuals

We can examine the residuals for bias and study their distribution. First, we consider the mean bias in the residuals for the seven indicators of Support for Tourism Development. The mean bias of the indicators ranges from 0.0006 to 0.0008 and thus shows very little mean bias. An inspection of the residuals charts will illustrate whether there is a skew in the residuals or whether they follow the standard normal distribution.

The predictive performance of the seven indicators of Support for Tourism Development are comparable and the residuals charts are highly similar. Therefore, we will consider only one residuals chart for the sake of brevity – refer to Fig. 5 for the residuals chart output from SmartPLS V3.2.7. RMSE is the first standard deviation of the predictive error and we have drawn the mean (solid line), the first (dashed line), and second (dotted line) standard deviations on this chart.

It is clear that the residuals are not normally distributed and there is a long left tail and a thick right tail. Residuals are calculated by subtracting the predicted value from the actual value ($y_{ij} - \hat{y}_{ij}$) and therefore negative residuals represent over-prediction and positive residuals represent under-prediction. It is interesting to note that there are a few extreme negative residuals, which warrant investigation.

5.4. Predictive Validity

Current software implementations of prediction for PLS-PM do well to generate OOS predictive metrics, but do not offer IS predictive metrics for the comparison of overfit. Therefore, generating IS metrics requires the use of additional spreadsheet software. To calculate IS metrics, we estimated the full dataset in SmartPLS V3.2.7 and exported the data descriptives (SD and mean)

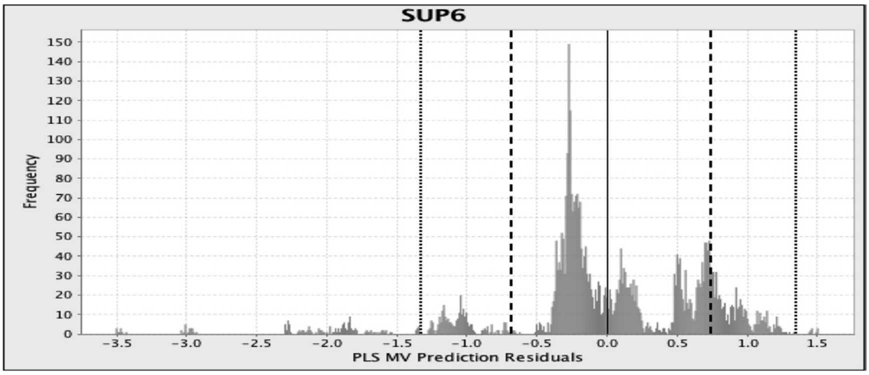

Fig. 5: Residuals Chart for SUP6. Solid Line is the Mean (0.0007), Dashed Line is the First SD (OOS RMSE: 0.706), and Dotted Line is Second SD (2 × OOS RMSE).

for outcome indicators, the original indicator scores, estimated outer loadings, estimated path coefficients, and estimated outer weights to a spreadsheet program. We then reproduced the technique outlined in Section 1 – we use the full dataset for both estimation and prediction and therefore generate IS predictions. Because the full dataset is used and IS prediction is desired, there is no need for cross validation. We use the full dataset for model training and prediction just as the cross-validation technique implemented by SmartPLS V3.2.7 generates OOS predictive metrics for the full dataset; thus, we can directly compare the metrics. Refer to Table 4 for the results.

The results of the comparison of OOS RMSE and IS RMSE demonstrate that there is very little loss in predictive power – less than a percent increase. Therefore, overfit is not a problem for this study.

5.5. Discussion of Empirical Example

The earlier analysis demonstrates that the PLS model analyzed generates predictions that are suitably better than the LM benchmark for the PP and Support for Tourism Development composites. The NP composite does not have suitable predictive power compared to the benchmark, but this is not surprising given the IS predictive power. The use of the predictive PLS algorithm is thus justified in this study.

The OOS predictive power of the model can be seen as fairly low in that the RMSE of the outcome indicators relative to the scale of the indicators is quite high. This could indicate that the model is suitable for general average-case predictions, but that care should be taken when practically applying these results in a managerial or practical context.

The mediators demonstrate a substantial increase in predictive power of the outcome indicators. In addition, this provides additional support for the

Table 4: Evaluation of Overfit for Support of Tourism Development.

Indicator	IS RMSE	OOS RMSE	OOS-IS RMSE	(OOS-IS)/ OOS (%)
SUP1	0.775	0.779	0.004	0.51
SUP2	0.747	0.751	0.004	0.53
SUP3	0.715	0.719	0.004	0.56
SUP4	0.708	0.711	0.003	0.42
SUP5	0.710	0.714	0.004	0.56
SUP6	0.702	0.706	0.004	0.57
SUP7	0.743	0.747	0.004	0.54

Notes. OOS-IS RMSE indicates the difference in predictive power of OOS and IS predictions for evaluation of overfit.

usefulness of the mediators as effective predictors of the outcome. The mediators are thus supported by the predictive analysis.

While there is little mean bias in the predictive error, the residuals charts demonstrate some issues. There are some cases, which suffer extreme negative predictive error. This demonstrates that the theoretical model is not suitably predicting these observations. The distribution is non-normal with two large peaks and many cases of overprediction.

Finally, the model does not display signs of overfit and can be said to display enough predictive power to predict new cases of the endogenous composite Support of Tourism Development.

6. Looking Ahead

While considerable progress has been made in the field of predictive PLS, there is still much that can be done. At the core of PLS-PM is the estimation of composite scores for unmeasurable latent constructs from measurable indicators. Therefore, the construct is the primary variable of interest to PLS-PM and techniques need to be developed and implemented that can evaluate predictive validity at the construct level.

Central to the evaluation of predictive validity in PLS modeling is the concept of IS versus OOS predictive performance – or, more specifically, overfit. However, there are still no software implementations of PLSpredict that allow us to explicitly evaluate overfit in the predictive context; nor guidelines as to how this evaluation could be conducted. Evaluation of overfit will add to the methodological rigor by providing an idea of generalizability of theory and allowing practical evaluations of the shortcomings of the model.

Further, if observations, their predictions, and predictive error are specifically identifiable, predictions could be used practically to isolate and identify specific cases where predictions deviate strongly from known observations. Cases with very poor predictive performance can be investigated to identify why they conform poorly to the model, potentially allowing for a follow-up survey or questions to be conducted and improving the theory.

We hope that this chapter has clarified the use of prediction in PLS-PM and provided you with specific actionable methods that can be applied to your research to complement evaluation of explanatory power with evaluation of predictive power and thus applying a joint explanatory and predictive approach.

Acknowledgments

The authors express their gratitude to S. Mostafa Rasoolimanesh, Mastura Jaafar, and Rabeeh Barghi (Universiti Sains Malaysia, Penang, Malaysia) for generously providing the dataset and domain expertise without which the empirical example would not be possible. The authors are also grateful for the invaluable feedback provided by Galit Shmueli (National Tsing Hua University, Hsinchu, Taiwan) and S. Mostafa Rasoolimanesh, which improved the overall quality of this chapter.

References

Chin, W. W. (2010). Bootstrap cross-validation indices for PLS path model assessment. In *Handbook of partial least squares. Springer handbooks of computational statistics*. Berlin, Germany: Springer Verlag.

Dijkstra, T. (2010). Latent variables and indices: Herman Wold's basic design and partial least squares. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields* (pp. 23–46). Berlin, Germany: Springer Verlag.

Evermann, J., & Tate, M. (2012). Comparing the predictive ability of PLS and covariance analysis. *Proceedings of the 33rd international conference on information systems*, Orlando, FL.

Evermann, J., & Tate, M. (2014). Comparing out-of-sample predictive ability of PLS, covariance, and regression models. *Proceedings of the 35th international conference on information systems*, Auckland, New Zealand.

Evermann, J., & Tate, M. (2016). Assessing the predictive performance of structural equation model estimators. *Journal of Business Research*, *69*(10), 4565–4582.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–152.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Berlin, Germany: Springer Verlag.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., … Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, *17*(2), 182–209.

Rasoolimanesh, S. M., Jaafar, M., & Barghi, R. (2017). Effects of motivation, knowledge and perceived power on residents' perceptions: Application of Weber's theory in World Heritage Site destinations. *International Journal of Tourism Research*, *19*(1), 68–79.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, *45*(5–6), 341–358.

Ringle, C. M., Wende, S., & Becker, J. M. (2015). SmartPLS 3. Bönningstedt: SmartPLS. Retrieved from http://www.smartpls.com

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, *69*(10), 3998–4010.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, *35*(3), 553–572.

Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research*, *69*(10), 4552–4564.