Association for Information Systems
AIS Electronic Library (AISeL)

7-1-2013

Modelling Mismatch In Predictive Analytics: A Case Study Illustration And Possible Remedy

Stefan Lessmann

University of Hamburg, Hamburg, Germany, stefan.lessmann@uni-hamburg.de

Follow this and additional works at: http://aisel.aisnet.org/ecis2013_cr

MODELLING MISMATCH IN PREDICTIVE ANALYTICS: A CASE STUDY ILLUSTRATION AND POSSIBLE REMEDY

Lessmann, Stefan, Institute of Information Systems, University of Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany, stefan.lessmann@uni-hamburg.de

Abstract

Predictive analytics (PA) is a popular approach to support managerial decision making. The core of a PA-based decision aids consists of an empirical prediction model. It is common to create such models using statistical standard methods. We argue that this approach suffers from a partial modelling mismatch (PMM) because the internal (statistical) objective of the decision support model and the business objective of the decision maker differ. We explore the severity and consequences of PMM in the context of resale price prediction, which is an important decision support task in the car leasing industry. The key hypothesis of the paper is that PMM decreases the quality of decision support. To test this, we develop a modelling methodology that creates predictive decision support models in a way so as to account for business objectives. Empirical experiments on real-world data confirm the effectiveness of our approach. In particular, we find evidence that i) PMM can substantially hurt decision quality, that ii) the new modelling approach is a suitable remedy, and that iii) it is generally important to consider the business objectives of decision makers when devising a predictive decision support model. With standard prediction methods falling short of this requirement, the paper makes a first step toward more business orientation in PA.

Keywords: Predictive Analytics, Decision Support Systems, Business Intelligence, Car Leasing.

1   Introduction

Predictive analytics (PA) is a major trend in the business intelligence landscape (e.g., Robb, 2012). In a decision support context, the role of PA is to generate model-based forecasts from observational data using a formal learning algorithm (e.g., Shmueli & Koppius, 2011). The goal of building an empirical model is to disclose information from large, high-dimensional data sets that would otherwise be inaccessible. PA features both a directive and non-directive decision support approach. Estimating the future consequences of alternative courses of action, PA forecasts inform decision making. In addition, a prediction model can act as a directive decision device that supersedes decision makers (Sutherland, 2008). Examples include online settings where a great number of decisions need to be made in very short time (e.g., Perlich et al., 2012) or settings where human judgement suffers cognitive biases and/or is vulnerable to opportunistic behaviour (e.g., Lessmann et al., 2010). The general theme of this paper is to contribute toward exploring the effectiveness of PA-based decision support.

The question how decision support systems (DSS) affect decision quality and, more generally, firm performance has been studied intensively in the IS literature (e.g., Lilien et al., 2004). The fit-appropriation-model integrates two streams of research that explain DSS effectiveness in terms of task-technology-fit and the way in which users appropriate a technology's features (Dennis et al., 2001). Related results have been obtained in the literature on forecasting support systems (e.g., Fildes & Goodwin, 2013). It is likely that findings on DSS effectiveness also apply to PA. More specifically, we conjecture that the fit of a PA-based decision aid affects its effectiveness. This is especially plausible in a directive framework where appropriation effects or users' ability of mitigate poor technology fit (Fuller & Dennis, 2009) are less important.

Relevant indicators to assess DSS effectiveness are profits, costs, or other business performance measures (e.g., Bharadwaj, 2000). Such measures also determine the effectiveness of PA-based decision aids. Intuitively, a prediction model that is meant to support decision making should consider the performance criterion that matters from a business standpoint. However, adopting this principle is not straightforward in PA. In particular, this paper studies a problem that we call partial modelling mismatch (PMM). While applications of PA are manifold and vary across industries, company sizes, and other factors, they share a common denominator. The support function follows from an empirical prediction model. PMM concerns the compatibility (or the lack thereof) between business objectives and the internal, statistical mechanisms that govern the development of such a prediction model.

A standard prediction method infers a functional relationship between a dependent variable and one or more independent variables from observational data (e.g., Hastie et al., 2009). To illustrate the problem of PMM, consider the case of linear regression. Fitting a regression model involves minimizing the sum of squared residuals over observations of the dependent and the independent variables. The sum-of-squared errors loss defines a 'model-internal' notion of performance in the sense that it governs the creation of the regression model. An optimal model is one with minimal squared-error. Other methods work alike. They also build prediction models through minimizing statistical loss functions (e.g., Hastie et al., 2009). In general, a prediction method and its internal loss function are closely connected. It is not straightforward to change one without changing the other.

Statistical loss functions are well grounded in theory and have many desirable properties. However, there is no guarantee that their notion of *performance* coincides with the business objective of a decision maker. Thus, there is room for a partial mismatch between the statistical objective, which is pursued during model building, and the business objective, which determines the model's eventual merit in a given application. Therefore, the goal of this paper is i) to examine how PMM affects the quality of PA-based decision support, and ii) to propose a modelling methodology that achieves a higher fit between modelling and business objectives.

It is important to note that our work on PMM is based on two assumptions. First, we assume that standard prediction methods are actually used for decision support. There is much literature to support this view (see, e.g., Bose & Xi, 2009; Ngai et al., 2011 for comprehensive literature surveys in specific domains). For example, standard prediction methods are used to target marketing campaigns (e.g., Verbeke et al., 2012), to inform risk management decisions (e.g., Baesens et al., 2003), and to forecast future sales (e.g, Lu et al., 2012), stock returns (e.g., Mettenheim & Breitner, 2010) or IS spending (e.g., Collopy et al., 1994). The prevalence of PA-based decision support in the literature gives a first proxy of its pervasiveness in industry. Furthermore, standard prediction methods are easily available in data mining packages such as SAS Enterprise Miner or IBM SPSS Modeler, which are widely used in industry (e.g., Gualtieri, 2013). Arguably, their core functionality is to solve standard data mining tasks including predictive modelling. Finally, several practitioners' conferences[1], white papers and case studies evidence the current interest in PA in industry (e.g., Robb, 2012; Vesset & Morris, 2011).

Second, we assume that PMM is a relevant problem in PA-based decision support. Put differently, even if PMM diminishes the suitability of standard prediction methods, one might expect that users find ways to mitigate poor technology fit (Fuller & Dennis, 2009). This is certainly true. One possible action involves developing a tailor-made modelling methodology that explicitly accounts for business objectives. There are ample examples of this approach in both the academic literature (e.g., Braun & Schweidel, 2011) and corporate practice (e.g., Kumar et al., 2008). However, self-developed solutions are notoriously more costly than using standard technology that is available in standard software. It is thus plausible that at least some companies adopt the latter approach and rely on standard methods. In this case, business objectives could be taken into at the model selection stage. That is, analysts could develop alternative models, simulate their performance in different scenarios, and then select the most suitable model for deployment (e.g., Verbeke et al., 2012). Such an approach introduces business objectives after developing the candidate models. The actual model building is still led by statistical loss functions, as there is typically no simple way to exchange a statistical loss function for a business objective (e.g., Hastie et al., 2009); especially when using standard data mining software. Given that options to circumvent PMM are scarce (i.e., when using standard prediction methods), we conjecture that PMM might be a relevant problem in several PA applications.

We pursue our research objectives in the context of a case study in car resale price forecasting. Resale price prediction is important to support decision making in the car leasing business (e.g., Pierce, 2012). As is explained in detail below, it is also a suitable context to explore PMM. Basically, this is because standard prediction methods assume symmetric costs of errors, whereas over-/underestimating car resale prices entail different costs. Hence, the modelling objective to minimize (squared) errors differs from the business objective to minimize costs. This allows us to examine whether PMM decreases the quality of PA-based decision aids and to test the effectiveness of the modelling approach that we propose to account for business objectives during model building.

In summary, the paper contributes to the theory and practice of PA-based decision support through i) calling attention to the problem of PMM, ii) illustrating its severity in the context of an important application by means of empirical experimentation on real-world data, and iii) developing a modelling methodology that achieves a higher fit with business objectives than standard prediction methods.

The remainder of the paper is organized as follows: Section 2 introduces the case study context and elaborates the problem of PMM in car resale price prediction. Section 3 describes our methodology. The experimental setup of the empirical study and corresponding results are described in Section 4 and Section 5, respectively. Section 6 concludes the paper.

---

[1] For example, http://www.predictiveanalyticsworld.com/.

2   Residual Value Estimation

Leasing has become an important way to market new cars. In a car leasing agreement, the lessee pays a monthly fee in return for receiving the right to use a car for a predefined period. The leasing rate covers depreciation and other costs related with offering the service. After contract expiration, the lessor is responsible for re-marketing the returned car in the second-hand marked. Estimating resale prices is crucial in this business because the leasing rate implies an expected residual value (e.g., Pierce, 2012). From the lessor's perspective, the leasing rate will be too high (low) if residual values are underestimated (overestimated). Offering high leasing rates above the market average will negatively impact sales performance. On the other hand, writing contracts with (too) low leasing rates diminishes returns and might lead to an overall loss from the lease when the car is resold below its expected residual value (e.g., Du et al., 2009).

In practice, it is challenging to measure the specific costs that result from inaccurate residual value predictions. Given the popularity of business intelligence, performance measurement, and related concepts in today's companies, it is however fair to assume that managers have some idea of the error costs of over- and underestimating resale prices. The key point of this paper is that there is little reason to believe that the error costs of under-/overestimating residual values are the same. For example, a risk-averse manager will be more concerned about actual losses (following from overestimating residual values) compared to opportunity costs from lost sales (following from charging high fees due to underestimating resale prices). A decision support model should be design such that it accounts for managers' preferences and requirements (e.g., Lilien et al., 2004). Accordingly, a resale price prediction model should strive to avoid the more costly type of forecast error.

Previous studies on residual value modelling do not reflect this principle. The prevailing approach is to estimate resale prices (as dependent variable) from car characteristics and possibly other factors (as independent variables) in a standard regression framework (e.g., Du et al., 2009; Erdem & Sentürk, 2009; Lessmann et al., 2010; Prado, 2009). This approach is formally given as:

$$y = f(\boldsymbol{x}, \boldsymbol{\beta}) + \varepsilon, \tag{1}$$

where y is the actual resale price of a used car, $\boldsymbol{x}$ a vector of independent variables, $\boldsymbol{\beta}$ a vector of model parameters that relate covariates to the response variable, and ε unperceived information (i.e., the model residual). After estimating $\boldsymbol{\beta}$ from data on past sales, the regression model $f(\boldsymbol{x}, \boldsymbol{\beta})$ facilitates predicting the resale prices of lease returns. More specifically, given that the lessor has full information on all car features, characteristics, etc. (i.e., the independent variables), s/he can use the model to predict a car's residual value before writing the lease. The regression function $f(\boldsymbol{x}, \boldsymbol{\beta})$ can thus be considered a decision support model that assists the lessor in setting prices.

Ordinary least squares regression assumes that the independent variables determine resale prices in a linear and additive manner. More advanced regression methods have also been explored (e.g., Lessmann et al., 2010). However, to the best of our knowledge, all previous studies create a regression model through minimizing a symmetric loss function such as the sum of squared residuals. Using a data set of n used-car sales, this approach is given as follows:

$$\boldsymbol{\beta}^* \leftarrow \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\beta}))^2. \tag{2}$$

The regression model is defined by the specific values of the model parameters $\boldsymbol{\beta}^*$. We argue that the approach to determine $\boldsymbol{\beta}^*$ through minimizing the sum of squared residuals is prone to PMM. The squared error function penalizes positive and negative residuals alike.[2] This way, the model building process does not account for differences in the severity of different error types. The optimal solution of the least-

---

[2] Note that positive and negative residuals are equivalent to underestimating and overestimating residual values, respectively.

squares minimization problem (2) and the optimal solution of a cost minimization problem in which residuals are weighted with their corresponding error costs will typically differ. This is especially true if positive and negative residuals (i.e., under-/overestimation) are associated with different costs (e.g., Granger, 1969). Consequently, the prediction models that follow from minimizing squared residuals and error costs, respectively, will also differ. Given that the economic consequences of under/overestimating resale prices are rather different (see above) a decision maker will be more concerned about one type of forecast errors. Positive and negative residuals are thus not equivalent from a decision maker perspective. But this is exactly the assumption of regression methods that employ symmetric loss functions. We hypothesise that creating resale price prediction models so as to account for asymmetric error costs will better support decision maker in setting profitable prices, and in this sense improve the quality of PA-based decision support. Subsequently, we describe the methodology that we propose to test our hypothesis and remedy the problem of PMM.

3   A Modelling Methodology to Account for Asymmetric Error Costs

A call for a deeper integration of decision makers' needs and preferences in PA will not spawn much dispute. Few would disagree that it is useful to concentrate on actual business users when developing decision support models. After all, actual business needs has always been central to the IS perspective toward decision support.[3] However, a note of caution is appropriate at this point. Predictive modelling is an induction problem: how should we build a model from a *sample* of observation such that the model generates accurate forecasts for novel cases not included in the sample? The key challenge is to create a prediction model that generalizes well to novel data; or, in formal terms, that achieves a low generalisation error. Established statistical loss functions embody much theoretical insight how model building should be organised to achieve this objective. Ignoring this insight and concentrating exclusively on business objectives such as cost minimization when building a prediction model is not a viable strategy. A model that concentrates exclusively on business objectives may give good performance on the sample of data from which it is created. But it will typically perform poorly on novel data. This is a well-known problem called over-fitting in statistical learning (Hastie et al., 2009). Therefore, we devise our modelling methodology such that it incorporates statistical and economic considerations.

It is difficult to formalize an actual business objective in car resale price prediction. Objectives differ across segments of the leasing market and companies (e.g., Pierce, 2012). Given that our main interest is to study PMM, we simplify this problem. In particular, we approximate business objectives with asymmetric cost functions (e.g., Granger, 1969). Such function, albeit being statistical in nature, embody the notion that different prediction errors are associated with different costs. This reflects the business context in resale price prediction and justifies our choice.

3.1   Asymmetric cost functions

A loss function measures the discrepancy between a forecast and the actual realization of the dependent variable. Examples are the absolute or squared difference between actuals and predictions. To assess prediction models, error measures aggregate the loss over a set of observations. For example, the well-known mean squared-error (MSE) is given as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 , \tag{3}$$

where $\hat{y}$ is a model-based forecast of the response variable *y*, and *n* the number of observations. Asymmetric cost functions extend symmetric error measures and are more appropriate when the costs of forecast errors differ across error types (Granger, 1969). Widely studied measures are the linear-linear

---

[3] Arguably, this is less true in other disciplines such as Operational Research that also aim at supporting decision making but put maybe more emphasize on the quantitative methods underneath a formal decision support system (see also Section 6).

(LLC), linear-exponential (LEC), and quadratic-quadratic (QQC) cost functions (e.g., Christoffersen & Diebold, 1997).

$$LLC = \begin{cases} \alpha|y - \hat{y}|, & \text{if } y > \hat{y} \\ b|y - \hat{y}|, & \text{if } y \leq \hat{y} \end{cases} \quad ; \quad QQC = \begin{cases} \alpha(y - \hat{y})^2, & \text{if } y > \hat{y} \\ b(y - \hat{y})^2, & \text{if } y \leq \hat{y} \end{cases} \quad ; \quad LEC = a\begin{bmatrix} \exp(b(y - \hat{y})) \\ -b(y - \hat{y}) - 1 \end{bmatrix} \quad (4)$$

The parameters *a* and *b* determine the severity of a particular error type. For example, a value *a<b* in LLC implies that overestimating *y* is more costly than underestimating *y*. To better illustrate the way in which errors are weighted, Figure 1 depicts the functional shape of the three cost functions (solid line) in comparison to the squared-error loss (dotted line) for different magnitudes of forecast errors.
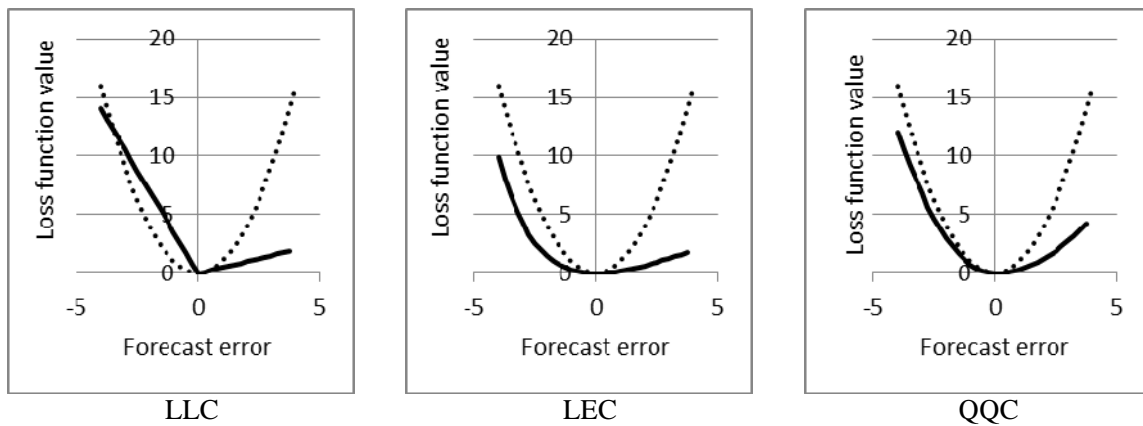

Figure 1.　　　*Asymmetric cost functions (solid line) versus squared-error loss (dotted line).*[4]

All three loss functions facilitate weighting different types of forecast errors according to their economic consequences. They are still crude approximations of business objectives in resale price predictions. However, they are better aligned with the pricing decision problem than the symmetric loss functions of standard regression methods. This suffices to examine whether increasing the fit between the modelling task and the decision task (i.e., the respective objectives) improves the quality of PA-based decision support. To test this proposition, we subsequently develop a methodology that can accommodate asymmetric loss functions during model building.

3.2　　Ensemble selection

An ensemble is a meta-model that includes several base models and combines their outputs to generate a composite prediction (e.g., Hastie et al., 2009). Ensemble selection (ES) is a two-step modelling approach. The first step involves creating a library of base models using standard prediction methods. Step two aims at identifying a beneficial combination of a sub-set of base models, using some search mechanism. The selected base models form the final ensemble.

We develop an ES approach on the basis of Caruana et al. (2006). Table 1 illustrates their strategy for a hypothetical library of *T* base models (M_1, M_2, … M_T). Panel 1 shows their forecasts together with actual resale prices for three cars. The current ensemble composition is shown in the rightmost column of Table 1. Caruana et al. (2006) begin with an empty ensemble. Using some loss function (MSE in Table 1), they assess all base models and add the best performing model (M_1) to the ensemble. Caruana et al. (2006) proceed with evaluating all possible combinations of the just selected base model and one other base model. This is shown in Panel 2. Consider for example M_2. Panel 1 shows its original predictions, whereas

---

[4] The values for *a* and *b* are 0,5 and 3,5 for LLC, 1,3 and -0,6 for LEC, and 0,3 and 0,75 for QQC. We select these settings because they produce loss values that are in a similar value range as the squared-error loss. This improves the readability of Figure 1. Note that our settings imply that overestimating resale prices is the more severe forecast error.

we compute the predictions shown in Panel 2 as a simple average over M_1 and M_2. Panel 2 reveals that combining M_1 and M_T decreases MSE compared to using M_1 alone and gives better performance than any other combination. Therefore, M_T is added to the ensemble. The evaluation of possible combinations of the current ensemble and one additional base model continues until ensemble performance stops to improve. In Table 1, adding a third model facilitates a further reduction of MSE (Panel 3), whereas the best four-base-model ensemble in Panel 4 performs worse than the previous ensemble. Thus, ensemble growing stops after four iterations with the final ensemble M_1, M_T, and M_1. Given that the ensemble prediction is simply the average over the chosen base models, multiple inclusions of the same model (e.g., M_1) give this model a higher weight. Hence, ES can switch between a simple and a weighted average-based combination rule (Caruana et al., 2006).

	y	M_1	M_2	…	M_T	Ensemble
Panel 1	59,05	60,47	60,52		69,22	M_1
	52,83	44,13	42,45		49,98	
	73,26	71,95	71,55		72,61	
	MSE	**26,47**	37,61		37,32	
Panel 2	59,05	60,47	60,495		64,845	M_1 & M_T
	52,83	44,13	43,29		47,055	
	73,26	71,95	71,75		72,28	
	MSE	26,47	31,79		**22,63**	
Panel 3	59,05	63,39	63,40		66,30	**M_1 & M_T & M_1**
	52,83	46,08	45,52		48,03	
	73,26	72,17	72,04		72,39	
	MSE	**21,85**	24,63		25,47	
Panel 4	59,05	62,66	62,67		64,85	M_1 & M_T & M_1 & M_1
	52,83	45,59	45,17		47,06	
	73,26	72,12	72,02		72,28	
	MSE	**22,24**	24,43		22,63	

Table 1.   Example of ensemble selection with a hypothetical base model library.

Most ES implementations use statistical loss functions such as MSE to select base models. We propose to reserve the second ES stage for economic considerations. More specifically, we propose to organize base model selection so as to optimize the performance of the ensemble in terms of a business performance measure, which reflects decision maker objectives. This approach allows us to address the two perspectives of PA-based decision support. The base models in the library rely on standard prediction methods and thus on established inference principles. The subsequent selection mechanism can choose only from the available – statistically valid – base models. Given this constraint, it seems safe to concentrate on the business problem and objective in the second ES step.

The base model selection strategy of Caruana et al. (2006) is well suited to test our modelling approach because it supports arbitrary performance measures. In this study, we use an asymmetric loss function for base model selection. We consider such a loss function a valid proxy of actual business objective in resale price prediction. It is at least closer to business objectives than a traditional symmetric loss function. Therefore, using an asymmetric loss function during model building (i.e., ES stage 2) achieves higher fit with the actual business problem and is a first step to overcome PMM.

4   Experimental Design

The data for the empirical study has been provided by a leading German car manufacturer. It comprises actual sales data for six different car models of the manufacturer's core brand. All car models belong to the premium or the middle-class segment of the market for passenger cars. Each transaction (i.e., a sale of one car in the second-hand market) is characterized by several attributes (age of the car, mileage, special features, etc.) and the resale price. Table 2 summarizes the data.

Car model	Cases	Number of independent variables for different variable categories							
		Age	Mileage	Customer	Model year	Engine type	Prev. usage	Colour / cushion	Special equip.
No. 1	124.386	2	1	12	6	25	10	44	78
No. 2	107.109	2	1	12	8	40	10	56	86
No. 3	132.504	2	1	12	7	16	10	40	86
No. 4	14.410	2	1	12	3	17	10	26	76
No. 5	70.624	2	1	12	5	24	10	36	81
No. 6	5.585	2	1	12	7	9	10	19	50

Table 2.   Characteristics of the car model data sets.

We build a library of prediction models for every data set (car model). The prediction models forecast resale prices from car attributes. We employ 13 different regression methods (Table 3) that represent a diverse set of alternative modelling paradigms (e.g., linear vs. nonlinear, parametric vs. nonparametric, single model vs. ensemble). Most techniques exhibit some metaparameters that allow the modeller to tune the method to a particular task. Examples include the number of hidden nodes in a neural network or the number of trees in a boosting ensemble (e.g., Lu et al., 2012; Ngai et al., 2011). We define a set of candidate values for such metaparameters and create one prediction model for every setting. This is a common approach to create a large library of diverse base models (e.g., Caruana et al., 2006). Table 3 gives the specific number of prediction models per regression method. Technical details of the regression methods are discussed in Hastie et al. (2009).

Regression method	Number of models
Multivariate linear regression	11
Stepwise linear regression	57
Ridge regression	39
Least absolute shrinkage and selection operator	39
Multivariate adaptive regression splines	60
Artificial neural network	126
Support vector regression	105
Regression trees	6
Nearest-neighbor regression	20
Bagged regression tree	9
Bagged neural networks	5
Boosted regression trees	11
Random regression forest	32
Total	**520**

Table 3.   Composition of the model libraries for each car model data set.

To evaluate alternative prediction models, we partition every data set into a training set (60%) and a hold-out test set (40%) using random sampling. The test set is never used during model building but reserved to assess the performance of fully-specified prediction models. This simulates a real-world situation where forecasts are generated for data that was not available when creating the model (e.g., Collopy et al., 1994). In addition, ES requires auxiliary validation data to assess base models and candidate ensembles in the second modelling stage (see Table 1). We obtain this data by 5-fold cross-validation on the training set (Caruana et al., 2006).

5   Empirical Results

One objective of this paper is to examine whether the effectiveness of predictive decision support models increases if business-oriented performance criteria are taken into account during model building. To test this in our application context, we compare ES with an asymmetric cost function as base model selection criterion to standard prediction models. In the following, we concentrate on the QQC cost function. An advantage of QQC is that it embodies the same quadratic error measurement as many symmetric loss functions in standard regression methods. This reduces the influence of other factors beside the symmetry/asymmetry of the loss function on the performance of alternative prediction models. In particular, we use QQC in the second ES stage (ES_{QQC} in the following) and also to assess the performance of different prediction models. This is because QQC is our proxy for actual business objectives in resale price prediction. In all experiments, a lower QQC value indicates better performance. We interpret better performance in QQC as an indication for improved decision quality.

We consider two benchmarks in our analysis. First, we compare ES_{QQC} to the best base model (BBM) in the library. This is a challenging benchmark because the BBM is selected from a large number of candidates including several state-of-the-art regression models (see Table 3). Furthermore, the BBM benchmark mimics a conventional PA approach where a standard prediction model is used without accounting for characteristics of the decision problem (e.g., asymmetric error costs). The second benchmark is based on the ES framework. Specifically, we create ensembles so as to minimize MSE during candidate selection. This approach (ES_{MSE}) differs from our proposed approach only in the strategy to select base models in ES step 2. Consequently, comparing ES_{QQC} and ES_{MSE} provides a clear view on the performance impact of incorporating a business performance measure into model building, compared to using statistical loss functions alone.

Figure 2 summarizes the results of the empirical evaluation. The ordinate shows the average (across the six car models) percentage difference in QQC between ES_{QQC} and BBM (dotted line), and ES_{QQC} and ES_{MSE}, (solid line) respectively. A positive difference indicates that ES_{QQC} performs better than the corresponding benchmark. Without loss of generality, we assume that overestimating resale prices is more severe than predicting a resale price below its actual value. This implies that the parameters *a* and *b* of the QQC loss function are such that *b>a* (see eq. 4). The ratio between these parameters determines the degree of asymmetry of the cost function: the smaller *b/a*, the more severe (costly) is an overestimation compared to an underestimation. To study the behaviour of ES_{QQC} in different scenarios, we fix *b* at one and consider ten settings of *a*=0,1; 0,2; …, 1,0. These setting represent different cost ratios of the form 1:*a*, and are shown on the abscissa in Figure 2.
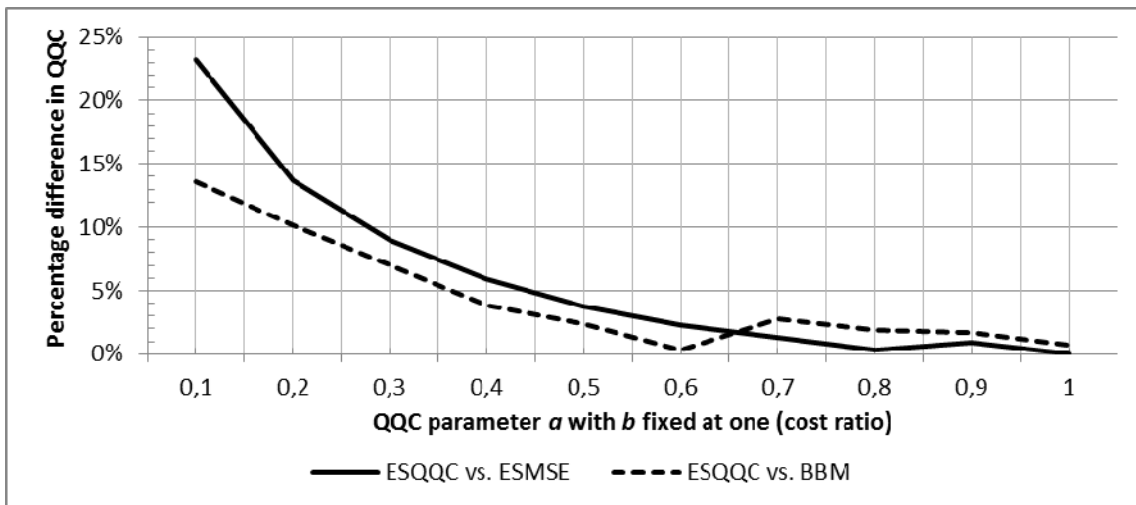

Figure 2.   *Percentage QQC difference between ES_{QQC} and two benchmarks for ten cost ratios.*

Figure 2 reveals that ES_{QQC} outperforms the two benchmarks in most settings, and performs never worse than BBM or ES_{MSE}. This is also confirmed by the results of a Wilcoxon signed rank test (Demšar, 2006). For both benchmarks, we can reject the null-hypothesis that the median performance of ES_{QQC} and the benchmark are the same (*p*-value ES_{QQC} vs. ES_{MSE} <0,000 ; *p*-value ES_{QQC} vs. BBM =0,0012). This evidences the merit of incorporating the performance measure that matters from a decision maker's point of view into the model building process. The approach that uses QQC for model building performs significantly better in terms of QQC.

The performance difference between ES_{QQC} and its benchmarks increases with decreasing values of *a*. This suggests that the degree to which ES_{QQC} improves upon a benchmark depends on the cost ratio between the two types of errors: the more skewed the error costs, the more important it is to account for the asymmetry during model building. This is plausible because heavily skewed error costs imply that the discrepancy between an ordinary symmetric loss function and QQC is large. From Figure 2, we may conclude that a standard loss function is less able to guide model building in such a setting. That is, the correlation between minimizing a symmetric loss function and minimizing QQC will be low. In the same way, the advantage of ES_{QQC} vanishes if different prediction errors are associated with similar costs. This performance trend offers further support for our choice to approximate business objectives with QQC. One may argue that QQC is still more similar to a standard loss function such as MSE than an actual business objective function (or utility-function) would be. Given that the performance gap between ES_{QQC} and, e.g., ES_{MSE} increases with the dissimilarity between QQC and MSE, the performance differences observed for QQC might be a conservative estimate for the performance differences that would follow from using an actual business objective within ES.

In summary, the observed results allow concluding that i) PMM exists in car resale price prediction, that ii) using an ES modelling framework with appropriately configured base model selection mechanism mitigates PMM, and that iii) such a modelling approach may improve performance substantially, if the discrepancy between a standard loss function and the business performance measure is large. The last conclusion follows from the results of Figure 2 at low values of *a*. A performance improvement of ten percent or more will be meaningful in many business applications.

6   Conclusions

We set out to examine PMM in PA-based decision support. The task of predicting car resale prices was used as case study context. We argued the suitability of this application, referring to the asymmetry of the error costs that follow from under-/overestimating resale prices. Drawing inspiration from the forecasting and machine learning literature, we developed a methodology that accounts for asymmetric error costs during model building. Empirical experiments evidenced the existence of PMM in car resale price prediction, and confirmed the effectiveness of our methodology. In particular, comparisons to challenging benchmarks showed that our approach performs significantly better than the competitors considered in the study. We also found that the improvement can be substantial if the actual business objective differs considerably from the performance notion of a conventional statistical loss functions. Overall, these results support the view that the quality of PA-based decision aids improves, if characteristics of the decision problem are accounted for when developing predictive decision support models.

Our study has practical and academic implications. From a practical point of view, the study calls attention to a shortcoming of standard prediction methods as decision support device. Business analysts are well advised to seek ways to increase the alignment between a decision problem and the representation of this problem in a formal prediction method. The study has also illustrated a possible way to achieve this. The ES framework advocates a modest extension of standard modelling practices. Many data mining packages provide a scripting environment to extend the base system's functionality. Implementing a base model selection strategy that maximizes a business objective might not be very difficult in such an environment. Examining how many decision tasks are supported with PA and how well statistical loss

functions approximate the corresponding business objectives should provide some guidance whether such an investment is worthwhile.

From an academic point of view, the study exemplifies a possible way in which the IS discipline can contribute toward future developments in PA, and more generally, data-centric decision support paradigms. Today, such concepts receive much attention among executives. Terms such as big data, data science, business/decision analytics or PA mirror this interest, and raise high expectations. Eventually, the success of any 'new' decision support concept will depend on the degree to which it increases decision quality and firm performance. The business context of a decision problem, managers' information, communication and collaboration needs, and the value proposition of IT-based decision aids have been studied for decades in the IS literature. This expertise is extremely valuable to scrutinize the effectiveness of PA as well as related analytics concepts, and to amend the quantitative methods underneath these concepts for more business orientation. This paper makes a first step into this direction through illustrating the problem of PMM and devising a possible remedy.

The study also suffers limitations that need to be addressed in future research. First, approximating business objectives with an asymmetric loss function is a simplification. In the face of diverse objectives across different segments of the car leasing market and different companies (e.g., Pierce, 2012), such a simplification was necessary for this study. However, it is important to examine the performance of predictive decision support models if actual business objectives are considered during model building. Second, all results have been obtained in the context of a specific application context using specific data from this application. PMM is a general issue. It is thus vital to gather more experience in other applications. This will help to better understand how, why, and under which conditions considering business objectives during prediction model building is effective. In this sense, a useful contribution of this study is that it provides a modelling framework that facilitates such tests.

References

Baesens, B., Setiono, R., Mues, C. and Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, 49 (3), 312-329.

Bharadwaj, A.S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. MIS Quarterly, 24 (1), 169-196.

Bose, I. and Xi, C. (2009). Quantitative models for direct marketing: A review from systems perspective. European Journal of Operational Research, 195 (1), 1-16.

Braun, M. and Schweidel, D.A. (2011). Modeling customer lifetimes with multiple causes of churn. Marketing Science, 30 (5), 881-902.

Caruana, R., Munson, A. and Niculescu-Mizil, A. (2006). Getting the Most Out of Ensemble Selection. In Proc. of the 6th Intern. Conf. on Data Mining, pp. 828-833, IEEE Computer Society, Los Alamitos.

Christoffersen, P.F. and Diebold, F.X. (1997). Optimal prediction under asymmetric loss. Econometric Theory, 13 (6), 808-817.

Collopy, F., Adya, M. and Armstrong, J.S. (1994). Principles for examining predictive validity: The case of Information Systems spending forecasts. Information Systems Research, 5 (2), 170-179.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7, 1-30.

Dennis, A.R., Wixom, B.H. and Vandenberg, R.J. (2001). Understanding fit and appropriation effects in group support systems via meta-analysis. MIS Quarterly, 25 (2), 167-193.

Du, J., Xie, L. and Schroeder, S. (2009). PIN Optimal distribution of auction vehicles system: Applying price forecasting, elasticity estimation, and genetic algorithms to used-vehicle distribution. Marketing Science, 28 (4), 637-644.

Erdem, C. and Sentürk, I. (2009). A hedonic analysis of used car prices in Turkey. International Journal of Economic Perspectives, 3 (2), 141-149.

Fildes, R. and Goodwin, P. (2013). Forecasting support systems: What we know, what we need to know. International Journal of Forecasting, 29 (2), 290-294.

Fuller, R.M. and Dennis, A.R. (2009). Does fit matter? The impact of task-technology fit and appropriation on team performance in repeated tasks. Information Systems Research, 20 (1), 2-17.

Granger, C.W.J. (1969). Prediction with a generalized cost of error function. Operational Research Quarterly, 20 (2), 199-207.

Gualtieri, M. (2013). The Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 2013. Forrester Research, Cambridge. http://www.forrester.com/pimages/rws/reprints/document/85601/oid/1-KWYFVB. Last accessed: 22.03.2013.

Hastie, T., Tibshirani, R. and Friedman, J.H. (2009). The Elements of Statistical Learning. 2nd ed. Springer, New York.

Kumar, V., Venkatesan, R., Bohling, T. and Beckmann, D. (2008). The power of CLV: Managing customer lifetime value at IBM. Marketing Science, 27 (4), 585-599.

Lessmann, S., Listiani, M. and Voß, S. (2010). Decision Support in Car Leasing: A Forecasting Model for Residual Value Estimation In Proc. of the Intern. Conf. on Information Systems (Lacity, M., Niederman, F. and March, S. Eds.), AIS, Paper 17.

Lilien, G.L., Rangaswamy, A., Van Bruggen, G.H. and Starke, K. (2004). DSS effectiveness in marketing resource allocation decisions: Reality vs. perception. Information Systems Research, 15 (3), 216-235.

Lu, C.-J., Lee, T.-S. and Lian, C.-M. (2012). Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. Decision Support Systems, 54 (1), 584-596.

Mettenheim, H.-J.v. and Breitner, M.H. (2010). Robust Decision Support Systems with Matrix Forecasts and Shared Layer Perceptrons for Finance and other Applications. In Proc. of the Intern. Conf. on Information Systems (Lacity, M., Niederman, F. and March, S. Eds.), AIS, Paper 83.

Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y. and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50 (3), 559-569.

Perlich, C., Dalessandro, B., Hook, R., Stitelman, O., Raeder, T. and Provost, F.J. (2012). Bid Optimizing and Inventory Scoring in Targeted Online Advertising. In Proc. of the 18th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining (Yang, Q., Agarwal, D. and Pei, J. Eds.), pp. 804-812, ACM, New York.

Pierce, L. (2012). Organizational structure and the limits of knowledge sharing: Incentive conflict and agency in car leasing. Management Science, 58 (6), 1106-1121.

Prado, S.M. (2009). The European used-car market at a glance: Hedonic resale price valuation in automotive leasing industry. Economics Bulletin, 29 (3), 2086-2099.

Robb, D. (2012). Gartner Taps Predictive Analytics as Next Big Business Intelligence Trend. http://www.enterpriseappstoday.com/business-intelligence/gartner-taps-predictive-analytics-as-next-big-business-intelligence-trend.html. Last accessed: 22.03.2013.

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

Sutherland, J.W. (2008). Directive decision devices: Reversing the locus of authority in human-computer associations. Technological Forecasting and Social Change, 75 (7), 1068-1089.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J. and Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. European Journal of Operational Research, 218 (1), 211-229.

Vesset, D. and Morris, H.D. (2011). The Business Value of Predictive Analytics. Working Paper, IDC. http://www.techrepublic.com/whitepapers/idc-whitepaper-the-business-value-of-predictive-analytics/5340922. Last accessed: 22.03.2013.