Decision Support for the Automotive Industry: Forecasting Residual Values using Artificial Neural Networks

Christoph Gleue[1], Dennis Eilers[1], Hans-Jörg von Mettenheim[1], Michael H. Breitner[1]

[1] Leibniz Universität Hannover, Institut für Wirtschaftsinformatik, Hannover, Germany
{gleue,eilers,mettenheim,breitner}@iwi.uni-hannover.de

Abstract. The leasing business is one of the most important distribution channels for the automotive industry. This implies that forecasting accurate residual values for the vehicles is a major factor for determining monthly leasing rates: Either a systematic overestimation or underestimation of future residual values can incur large potential losses in resale value or, respectively, competitive disadvantages. In this paper, an operative DSS with the purpose of facilitating residual value related management decisions is introduced, with a focus on its forecasting capabilities. Practical implications are discussed, a multi-variate linear model and an artificial neural network approach are benchmarked and further, the effects of price trends and seasonal influences are investigated. The analysis is based on more than 150,000 data sets from a major German car manufacturer. We show that artificial neural network ensembles with only a few input variables are capable of achieving a significant improvement in forecasting accuracy.

Keywords: Decision Support Systems, Business Intelligence, Artificial Neural Networks, Residual Value Forecasts, Car Leasing

1 Introduction and Motivation

The leasing market is an important business segment for car manufacturers. Automotive assets, which include passenger cars and commercial vehicles, accounted for a volume of 65% (178.2 billion Euros) of total new leasing contracts granted in 2014, which makes it the largest individual asset segment of the European leasing market [1]. Almost a third of all new cars sold by German premium brands in 2015 was financed with leasing [2]. Hence, this business model provides tremendous market opportunities, particularly for car manufacturers and leasing companies. Nevertheless, there are also risks which are difficult to quantify. The focus of our research is the so-called residual value risk [3]. While the customer pays a contractually-fixed leasing rate over the entire period of use, the residual value of the vehicle is uncertain until the end of the leasing period. In order to compensate the loss in value of the vehicles with adequate leasing payments, forecasts of residual values must be as exact as possible. Hence, accurate forecasts of future residual values constitute a critical success factor for competitiveness in the leasing market. Either a systematic over-estimation or under-

13th International Conference on Wirtschaftsinformatik,
February 12-15, 2017, St. Gallen, Switzerland

Gleue, C.; Eilers, D.; von Mettenheim, H.-J.; Breitner, M. H. (2017): Decision Support for the Automotive Industry: Forecasting Residual Values using Artificial Neural Networks, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 1205-1219

estimation of future residual values would have negative consequences. If an over-estimate is determined, the leasing payments may be lower, but not sufficient to compensate the loss in value of the vehicle. If predicted residual values are systematically too low, higher leasing rates must be set. This means that either avoidable competitive disadvantages occur or leasing rates must be substantially subsidized by car manufacturers. Car manufacturers nowadays often already have a sufficiently good data history of completed transactions in the leasing sector. They also record resale prices in the used car market for the corresponding vehicles. This potential, however, is often not sufficiently utilized to improve predictions of residual values.

In our research, we introduce an operational Decision Support System (DSS) which was developed and successfully implemented in cooperation with a large German car manufacturer. Although the incorporated higher-level analyses, management reports and visualizations are equally important parts of the aforementioned DSS, they all rely on proper data handling and a reliable forecasting methodology. Thus, in this paper we focus on the forecasting module of the system. The residual value forecast is a typical regression problem in which certain characteristics such as the age and mileage of a particular vehicle determine its residual value. The study is based on more than 150,000 records of completed leasing transactions and resale values over a period of four years (2011-2015). Linear models and Artificial Neural Networks (ANN) are implemented in this study. For this purpose, the development of vehicle values depending on the model age and time factor is of particular interest. The model age measures the time between the market launch of a new model class and the resale date of a specific vehicle belonging to this class. The continuously increasing model age means that the expected residual value of two identical vehicles of exactly the same age and with the same mileage is not constant over time. In addition, prices are influenced by general market conditions, accounted for by a time component in our analysis. Furthermore, the residual values on the used car market are subject to seasonal fluctuations within a year (a critical factor for ANN forecasts [4]). In order to allow for an unbiased forecast of estimated residual values over several years, these influences are examined more closely.

The remainder of this paper is structured as follows. The next section 2 provides an overview of the existing literature in the field of residual value forecasts. The structure of the DSS and the available input data are presented and explained in section 3. Section 4 specifies an exploratory analysis about trends and seasons in the used car market and subsequently introduces the forecasting model based on ANNs. The results of the analysis are presented in section 5. Section 6 provides a discussion and addresses limitations, implications for practical use and further research opportunities. Section 7 concludes with a short summary.

2 Related Work

This section presents the existing literature in the field of residual value forecasts. Due to the exclusivity of the data, only few studies on this subject exist. This fact, however, also highlights the enormous research potential. One of these studies is provided by Lessmann et al. [5]. On the basis of 124,386 transaction data of the same vehicle model (upper class) of a major car manufacturer, the authors develop a decision support system by means of a support vector regression (SVR). This method extends the classical linear regression such as to permit a non-linear transformation of the independent variable. Each vehicle is described by 176 attributes. The large number of attributes results from the use of dummy variables that represent features such as different optional equipment. Transaction-specific characteristics such as typical features of the customer constitute a crucial point in this method. The authors demonstrate the benefits of using this information in a forecasting model. As such, they recommend the expansion of residual value forecasts within the company to achieve improved predictive power on the basis of exclusive datasets which are not available to external service providers or residual value institutes.

Wu et al. [6] forecast used car prices on the Taiwanese market. Their input parameters comprise the cars brand name, the year of manufacture, the engine type and an equipment index. A new combination of ANNs and ANFIS (adaptive neuro-fuzzy inference system) models are proposed to improve forecasting accuracy. An earlier study carried out by Lian et al. in 2003 describes the problem of the residual value forecast from a time series perspective [7]. Evolutionary Artificial Neural Networks (EANNs) are used in this study to model the residual value of vehicles (all 24 months old) over time (from 1993 to 1997). The authors find cyclical fluctuations, according to which the residual value is at a high level at the beginning of the year and falls to a lower level towards the end of the year (see section 4.1 of our paper for a similar seasonality analysis). Other studies often use mostly macroeconomic indicators in addition to internally available data sources in order to explain the residual value distribution. Prado implements the price of diesel fuel and the industrial production index as explanatory variables in addition to vehicle-specific variables such as age and mileage [8]. At the present time, it is evident that no scientific standard methods are reported in the respective literature. Fan et al. provide a comparison between data mining model approaches such as AutoRegressive Trees (CART), ANNs and linear regression [9]. Their investigations focus on heavy construction machines. According to their analyses, the CART model provides the best results compared to the ANN and the linear regression model.

Besides the purely data-driven forecast techniques, other theoretical model approaches exist to explain price developments and the implications for risk management. This aspect is not a focal point of our work. For a deeper insight, we refer to [10], [11] and [12].

3 System Structure and Input Data

In the following sections, the data and methodology presented in this paper are put in the larger context of a DSS for residual value- and leasing-related issues. In section 3.1, an overview of the complete system structure (see figure 1) is presented, followed by a more in-depth description of the input data used for this study in section 3.2.

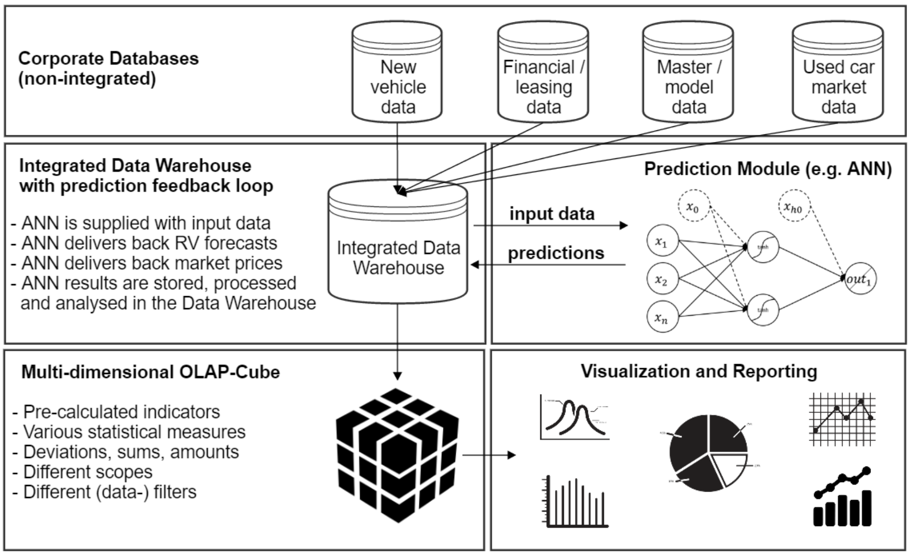

Figure 1. Decision Support System for RV forecasts

3.1 Structure of a DSS for Residual Value Forecasts

In this section, the larger context of the DSS in which the prediction module we focus on is implemented, is explained (see figure 1). The forecasting models need to be trained with historical data. Thus, the first step of the development was to collect and integrate data from different sources within the corporation. The most important of all sources is the used car market database, which stores all used car data collected from the dealerships. It contains every single transacted resale in Germany since 2011, including the Vehicle Identification Number (VIN) as a unique identifier for the vehicle, its mileage, the age of the vehicle, additional model information (e.g. model type, engine, fuel type etc.), resale date and the desired output of the forecasting model, the net resale price. As all data from this source is entered manually in the dealerships and then transferred via a common interface for all dealer management systems in the market, it has to be joined and validated with data from other, internal sources to serve as training data set for the forecasting model (see section 4.2).

One of those sources is the new vehicle database, which collects data from the factory invoicing system directly after production. Besides the corresponding VIN numbers, it contains the original list price of the vehicles, divided into base price, color surcharge and the extra charge for optional equipment, the complete list of optional equipment packages and also an indicator if the vehicle was produced for a promotional offer (with a discount on the original list price).

The information delivered by these systems is decoded using additional data from a Master Data Management system which contains general information about model types, promotional campaigns (such as list price discounts on special edition models), engine power and capacity, gearbox type (manual, automatic, double clutch), number of gears and the model age (number of months since the market launch). Ultimately, there is a connection to the leasing database which provides the initial forecasts and market prices assumed by the captive leasing company. For every single resold vehicle, all of the aforementioned information is gathered, cleansed and stored automatically in a central, integrated data warehouse. This data warehouse, in turn, has a direct connection to the prediction module. The (trained) forecasting model is being fed with input data directly from the data warehouse, predicts market values (actuals) and expected residual values for all vehicles in the database and delivers them back to the data warehouse, where they are stored for further processing and analyses. The next and penultimate layer of the DSS is a multi-dimensional cube for Online Analytical Processing (OLAP). It serves as the basis for visualizations and reports for different purposes, e.g. benchmarking of the forecasting models, market analyses and, most importantly, leasing- and pricing-related decision support. It provides numerous pre-calculated indicators, deviations, sums, amounts, arithmetic means and other statistical measures as well as pre-configured scopes and data filters, e.g. to differentiate between the list price of special edition and standard models or young used cars and cars representing the traditional leasing segment with a leasing term of at least 12 months.

Hence, the OLAP cube can be seen as an extensive toolbox for management reports and visualizations. Used correctly, it can be utilized to find answers for many different used-car- or leasing-related questions, as e.g. individual dealership sales performance rankings, analyses of the used car market price levels after external shocks, monitoring of the car manufacturers' own residual value setting or influences of particular parts of optional equipment on the resale price. These reports need to be configured once on the OLAP level and can then be standardized and visualized with additional software.

The modules previously described, from data collection and integration to reporting and visualization, all play an important role in the operational DSS outlined in this section. Nevertheless, all the results and reports in our particular area of application rely on accurate predictions of actual market values and, even more importantly, most accurate residual value forecasts. Subsequently, the prediction module and its mode of operation represent the core of our DSS and thus, the main focus of the research presented in this paper. The next section 3.2 provides a description of the input data used for the forecasting model.

3.2 Input Data

The data covers a period of four years from 2011 to 2015. Vehicles are included in the database as soon as they have been successfully resold on the used car market after the end of the leasing period. By this means, the residual values realized on the market are included in the database. We distinguish between individual vehicle-specific and model-specific variables. The vehicle-specific variables refer to the leasing contract agreements, which include the mileage, the leasing term (age of the vehicle) and extra charges for special colors and optional equipment. The model-specific variables describe the general characteristics of the vehicle and are the same within each model group. 928 different models are represented in the data. Table 1 shows the available vehicles' features.

Table 1. Input data variables

Variable	*Type*	*Description*
pricecolor	continuous	Extra charge for the color as a percentage of the list price
priceequipment	continuous	Extra charge for optional equipment as a percentage of the list price
mileage	continuous	Vehicle mileage in km/100,000
agevehicle	continuous	Vehicle age (registration date to resale date in days)
pricecolor	continuous	Extra charge for the color as a percentage of the list price
agemodel	continuous	Age of the model since its market launch in months
enginecapacity	continuous	Engine capacity in cubic centimeters
enginepower	continuous	Engine power in horsepower
fourwheel	binary	Four-wheel drive indicator
fuelcode	binary	Gasoline (B), Diesel (D), CNG (EG), LPG (AG)
geartype	binary	Transmission types: automatic (A), double-clutch (D), manual (S)
residualvalue	continuous	Residual value in percent ($\frac{resale\ price}{list\ price}$)

These variables are directly related to the vehicles under investigation. Using the available data, a linear regression analysis is first carried out to explain the residual values. The linear model is specified with all the variables listed in table 1. As expected, the age of the vehicle as well as the mileage have a significant negative impact on the residual value. For a forecasting application, the model age is also a crucial factor. Here we face the problem that the age of a specific vehicle model constantly increases over time. Compared to the age of a specific vehicle where we have a wide range of training examples, the age of the vehicle model is not available beyond the present point in time. The results of this regression analysis indicate that in-sample, the effect of the model age is significantly negative. Using this factor in a forecasting application therefore requires extrapolation techniques. Another interesting finding is that the extra charge for optional equipment has a negative effect on the residual value, which shows that the more optional equipment a vehicle is fitted with, the more value it loses over time in percentage terms.

4 Forecasting Model Specification

4.1 Trend and Seasons in the Used Car Market

In this section we use the residuals of the previously mentioned regression analysis to investigate the phenomenon of seasonality and trend in the used car market, which provides valuable insights to perform a subsequent unbiased forecast. Using the regression approach, all effects of the aforementioned independent variables are eliminated. The following approach is applied: firstly, residuals are sorted with respect to time. Secondly, the residuals for each month during the observation period are averaged. This results in a monthly time series of averaged residuals. Figure 2 shows the resulting time series.

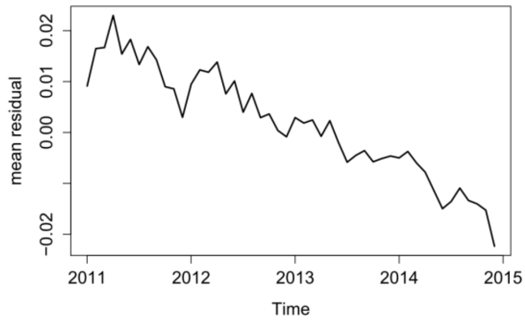

Figure 2. The influence of the time factor on the residual value

The results show a clear downward trend during the four years of observation. After controlling for all contract-specific variables, the results indicate a price decline based on general market conditions (the regression already controls for the constantly increasing model age) for used commercial vehicles. Since we are also interested in seasonal patterns, a decomposition of this time series is performed by the Seasonal and Trend decomposition using Loess (STL) method [13]. A seasonal period of 12 (for each month) is assumed. Figure 3 shows the time series decomposition to determine seasonal and trend effects. The ordinate shows the mean residuals in total (data) and their three different components. It can be observed that the residuals are subject to a trend and a season. The detrended seasonal component shows that towards the end of the year, lower residual values are achieved than during the rest of the year. Accordingly, the residual values tend to be higher during spring, even though these effects are rather small (in the range of 0.8 percentage points of the list price during a year). In order to control for the effects analyzed above, a new regression model is specified according to equation 1, which incorporates a vehicle model independent time factor and also controls for the seasonal components using monthly dummy variables.

$$rv_k = c + \sum_{j=1}^{13} (\beta_j \cdot feature_{j,k}) + \gamma \cdot Time_k + \sum_{m=1}^{11} (\delta_m \cdot Month_{m,k}) + \varepsilon_k \qquad (1)$$

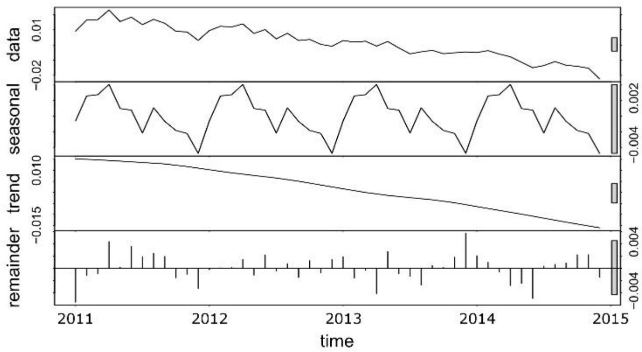

Figure 3. Time series decomposition: seasonal and trend effects

We use the resulting residuals to test for misspecification due to non-linearity with the BDS (Broock, Dechert and Scheinkman) test [14]. The null hypothesis (ε_k is iid) is rejected by the test at the 1% level ($\Delta\varepsilon = \sigma\varepsilon$ and embedding dimension m = 2). Therefore, the linear specification is still rejected by the BDS test. In a forecasting application which can be used in practice, there are tools available to incorporate non-linear relationships in a very effective way. These require careful data handling in comparison to linear models to avoid misleading results. This especially applies to noisy real world data. ANNs are introduced in the following section, which also provides a description of how to preprocess and prepare the data on hand.

4.2 Forecasting with Artificial Neural Networks

ANNs [15] are particularly suitable for the use with non-linear relationships. Their applications range from forecasts in the finance [16], [17] and risk management [18] area to Decision Support Systems [19], [20]. They are robust to very noisy, unstructured, or missing data [21], have powerful pattern recognition capabilities [22] and are therefore well suited to the present real-life problem. Even though many previous studies have shown the potential of this method in forecasting and prediction applications, a proper implementation and validation as well as an accurate data pre-processing is necessary to achieve reliable results [23]. ANNs may be viewed as a method for non-linear function approximation. In this paper, feed forward networks are used. These are composed of several layers of neurons. A first layer (input layer) describes the independent variables which are used to explain the phenomenon. These neurons are connected by weights θ_1 to a further layer (hidden layer), which in turn is connected to the output $h_\theta(X)$ (or a further hidden layer) by the weights θ_2. The hidden layer is thus in turn the input for the following layer. This pattern can be repeated any number of times (any number of hidden layers), each with any number of neurons within the layers. A three-layer feed forward ANN is defined by:

$$h_\theta(X) = \theta_2 \tanh(\theta_1 X) \tag{2}$$

The hidden neurons and, optionally, the output neurons transform their weighted sum of inputs by means of an activation function (usually the hyperbolic tangent).

$$f(X) = \tanh(X) = 1 - \frac{2}{e^{2X}+1} \tag{3}$$

The structure of an ANN is shown in figure 4. To be able to perform a function approximation, the ANNs are trained with training patterns (input x_k, output y_k represent the training pattern k). The randomly initialized parameters (weights) are determined using an iterative process. To avoid the problem of overfitting, a cross-validation is performed. Therefore, the training patterns are split into a set on which the ANNs are actually trained, I_t, and a set of validation patterns, I_v, to estimate the out-of-sample performance. The approximation quality of the ANN is evaluated by calculating the training and validation error functions

$$\varepsilon_t \equiv \frac{1}{2}\sum_{k\in I_t}(h_\theta(x_k) - y_k)^2 \text{ and } \varepsilon_v \equiv \frac{1}{2}\sum_{k\in I_v}(h_\theta(x_k) - y_k)^2. \tag{4}$$

We use an early stopping approach (error on the validation set increases) to obtain ANNs with good generalization capabilities. The best performing ANNs (based on validation error) are then combined in an ANN ensemble, so that the arithmetic mean is used as the actual result of the forecasting model. In this study, the “Fast Approximation performed with Universal Neural Networks” (FAUN) neurosimulator [24] is used for this purpose.

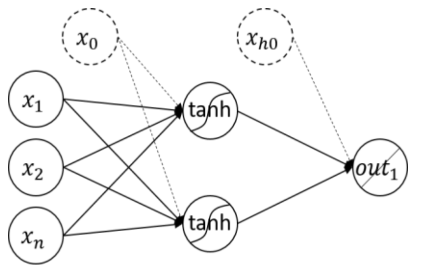

Figure 4. ANN (three-layer feedforward network)

As mentioned above, a usable forecasting application involves constantly increasing, time-dependent parameters: the model age and the time factor in general, which includes all external economic influences and market conditions. Both factors and their influence are not measurable (represented in the data) beyond the present point in time. Including data from the whole time period in the training process (e.g. separating training and testing data randomly from the database) leads to a look-ahead bias in the forecasting result. The training and forecasting periods are presented in figure 5. The forecasting model is trained with data from 2011 and 2012. The out-of-sample performance is documented with data of completed leasing contracts in the following two years (half-year intervals).

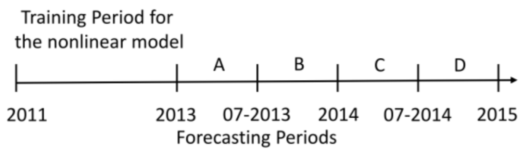

Figure 5. Forecast time line

Since ANNs are sensitive to outliers, we perform a data cleansing before the actual training process is started. For this purpose, 2000 three layer perceptrons including all variables shown in table 1 are trained to a local minimum on the training data. The best ANN (in-sample) is selected and tested for errors. Subsequently, every data set for which $|h_\theta(x_k) - y_k|$ is greater than a threshold value c is removed. As the threshold value c is defined as the 95% quantile of the errors, this means that 5% of the data are removed. This process is repeated a second time. We choose this approach as a result of an extensive cross validation process in which we evaluate the out-of-sample errors of the resulting ANN models with different data cleansing methods (e.g. removing more/less data; performing the procedure only once or three times; manual outlier removal). A manual analysis of the identified outliers reveals that in the case of outliers with higher residual values than expected, additional equipment (like e.g. additional cooling or tool storage equipment) had been installed after specifying the leasing contract, which led to a higher initial value and thus, also to a higher residual value. In the case of outliers with lower residual values than expected, the cars' overall condition (e.g. accidents) had not been properly reported in the data filtered by this approach. As we aim for accurate residual value predictions for accident-free vehicles without additional installations, these outliers had to be excluded from our analyses. For the actual training of the forecasting model, three-layer perceptrons with 4, 6 or 8 hidden neurons were tested. For each of the different topologies, 2000 networks were randomly initialized and trained to a local minimum. On the basis of validation data which is not represented in the training data set, the best 30 networks were selected and the arithmetic means of these approximations were used as a result. The ANN models with 6 hidden neurons yielded the smallest error regarding the validation data. In the next section, the results of the analyses are presented.

5 Results

This section presents the results of the linear and non-linear function approximations. ANNs generally provide good results for interpolation tasks. This means that the function approximation works well if the presented values lie within the input/output ranges of the training data. An extrapolation far beyond these ranges is difficult to realize. The forecasting application with ANNs for the third and fourth year (hence not represented in the training data regarding the time variables) uses the characteristics of new vehicles as inputs but sets the model age and time factor to the maximum value represented in the available data for these periods. The result is a so-called current market value, which reflects actual prices on the used car market. Table 2 (a) shows the resulting bias (mean error defined in equation 5) when using these forecasting models for predictions in the following two years.

$$mean\ error = \frac{1}{n}\sum_{i=1}^{n}(real\ value_i - forecast_i) \quad (5)$$

Table 2. Adjusted and non-adjusted forecasts of the ANN

(a) Non-adjusted forecast

Period	*Mean error*	*P-value*
A	0.00836	0.0524
B	-0.01028	0.0074**
C	-0.01551	<0.001***
D	-0.04640	<0.001***

(b) Adjusted forecast

Period	*Mean error*	*P-value*
A	0.00836	0.0524
B	0.00220	0.5637
C	0.00533	0.1817
D	-0.01787	<0.001***

***, **, * indicate statistical significance at the 0.1%, 1% or 5% level, respectively.

A t-test with the null hypothesis that the mean error is zero (unbiased) already rejects the hypothesis in the second forecasting period at the one percent significance level.

The actual residual value (forecast), in contrast to the market value, is now adjusted by a linear factor resulting from a time series regression of the first two years of data. The adjustment of the ANN forecast is performed according to equation 6.

$$rv_{t+n} = Market\ Value_t + \left(\beta_{agemodel} + \beta_{Time}\right) \cdot n \tag{6}$$

Table 2 (b) shows the results after the linear adjustment. The adjusted results show that the bias is now also insignificant for the forecasting periods B and C. For period D, the forecasting model is still biased. An investigation of possible non-linear dependencies in the time dimension is a topic for future research. This implies that a database of more than one complete business cycle must be investigated. The linear adjustment currently undertaken is a correct assumption for at least one and a half year forecasts, as indicated by the time series decomposition shown in figure 3 and the results listed in table 2 (b). The proposed linear model specified by equation 1 is now tested against the ANN forecast with a linear adjustment for the model age and time factor. To evaluate the linear model in a forecasting benchmark, a regularization term is incorporated for the training process. We choose a ridge regression approach (L2 norm) to prevent potential overfitting of the linear model. The lambda parameter is selected by cross validation. Due to the small number of parameters compared to the large number of training patterns in the years 2011 and 2012, the lambda parameter with the best performance is close to zero (0.04), so the linear model barely suffers from a high variance problem. Table 3 shows the results, the root-mean square error (RMSE) and the mean-absolute error (MAE), of the two forecasting methods for each of the four time periods. As evident from the increasing trend in the error measures, the forecast becomes more inaccurate the longer the forecasting period is. The non-linear relationships between the vehicle characteristics are represented significantly better by an adjusted ANN method than by a pure linear model. The Diebold-Mariano test [25] rejects the hypothesis that both methods have the same accuracy at any confidence level.

Table 3. Benchmark results

Period	*Measure*	*Adjusted ANN*	*Linear*	*DM-Test*
A	RMSE	0,07834	0.08554	<0.001***
	MAE	0.05638	0.06287	
B	RMSE	0.08026	0.08824	<0.001***
	MAE	0.05821	0.06515	
C	RMSE	0.07995	0.08965	<0.001***
	MAE	0.05891	0.06788	
D	RMSE	0.08218	0.09113	<0.001***
	MAE	0.06051	0.06857	

***, **, * indicate statistical significance at the 0.1%, 1% or 5% level, respectively.

Besides the significant performance improvements, these results are indeed economically relevant in our practical application compared to former, simplistic predictions based on data from external service providers for residual value estimations. In the next section, we discuss the results and provide an outlook for further research.

6 Discussion

The challenge of forecasting residual values for commercial vehicles was examined in this paper. An accurate forecast as a substantial part of a larger DSS is crucial for managing the exposure risk of car manufacturers and dealers in the leasing business. A systematic bias in both directions (too high or too low predictions) leads to negative consequences, either for competitiveness or the resale margin. The aim was to establish a purely data-driven approach for forecasting residual values based on information about past transactions. The results indicate that ANNs are well-suited for such noisy and unstructured data. Nevertheless, thorough data preparation and outlier detection is important in order to achieve good results. We show how ANNs can be used to preprocess and filter the database and how a forecasting model can be designed based on an ANN approach. ANNs are often criticized owing to their black-box nature. By means of a transparent description of the topology selection and data cleansing process, it is possible to mitigate the problem and make the results reproducible. ANNs are often only used at present as an alternative tool to benchmark different methods. However, an optimization within the whole ANN framework can help to improve the forecasting capability of this method.

Accordingly, the influence of the model age on the residual value over time and a general time factor have been investigated in more detail. This is necessary due to the fact that the model age (the period of time a specific vehicle model has been on the market since its market launch) is a factor which is not measurable beyond the present point in time, as it is a continuously increasing value. This also applies to the time factor, which takes account of all external market conditions. Although both factors remain uncertain, they play a crucial role in an actual, usable forecasting application. All other explanatory variables such as the age of a specific vehicle or its mileage are represented in the data for a sufficiently wide range of values. In the present study, the time-dependent factors were investigated in detail in order to linearly adjust the non-linear model to mitigate this problem. A clear downward trend in the residual values was identified with increasing model age and time. A reason for this may be the declining attractiveness of a particular vehicle model, which means that customers tend to buy a newer generation of products. As a result, older vehicle models must be offered at a lower price.

Moreover, the seasonal component of the used car market shows systematically lower residual values in December and higher residual values during the first quarter of a year. This empirical observation may be explained, e.g. by dealers trying to meet their sales targets towards the end of the year. Although these observations might be a special phenomenon for the type of vehicles examined in our study (commercial vehicles), it is a topic for further research investigating these patterns for a broader range of vehicle classes in order to formulate a more general statement. The entire analysis carried out in this paper is based on data between 2011 and 2015. From an economic point of view and in relation to the automotive market, this period does not include any major crises or boom phases. As the DSS and all necessary interfaces presented in section 3 are already successfully implemented, market data, as well as the corresponding internal data from corporate systems and databases, is being automatically recorded and stored in the central data warehouse for further processing. It is of interest to investigate how the results and the forecasting model quality change by analyzing data covering a longer period of time, including information about the behavior of residual values during crisis and/or boom phases in the database.

As previously mentioned, it is possible that external factors such as consumer demand, competitors' decisions, fuel prices or macroeconomic developments, e.g. measured by an industrial production index, also have a significant impact on the achievable resale price. Such factors were not included in this study for two reasons. Firstly, a complete theory which explains all factors that influence the used car market does not exist. This could take the form of a "fishing license" (alluding to the factor models of asset pricing). Up to now, however, not enough data is available to test the out-of-sample influences with a sufficiently long time horizon. Secondly, the problem of a look-ahead bias is the major concern in this application. The factors which drive prices on the used car market must be measurable at the time of the contract conclusion. This is a necessary precondition for performing a usable forecast in practice. In order to avoid biased predictions, however, economic indicators must also be forecasted to provide additional forecast model input data. To achieve a broader understanding of residual value data in general, an extensive study about different kinds of internal and external factors must be conducted in addition to benchmarks of proposed forecasting methods in the literature. Since each of the aforementioned, related studies have their own protected data with different characteristics, a benchmark of their methods is hardly interpretable if it is performed solely on one specific data set, because one method may not suit all kinds of data. Addressing this issue is an important topic for further research.

The above paragraph of the discussion is directly focused on the forecasting methodology implemented in our system. In the greater context of an entire DSS, the particular use case presented in this paper serves as a general example that by using methods and tools from the field of business intelligence, predictive analytics and data science, corporations can obtain valuable, business-critical information from data that is often already internally available. It also shows that this information, properly prepared and visualized, can support or even induce management decisions and help to monitor the consequences. In the particular case of the DSS presented in section 3, many use cases beyond residual value forecasts have already been realized, from the assessment of individual dealership performance to the measurement of the impact of internal decisions or external effects. Another practical example from our project is the simulation of different market scenarios, e.g. the assessment of the car manufacturers' risk exposure if residual values suddenly drop due to an unexpected external shock such as financial crises, new legislations or competitors' decisions. The results of such "stress tests" support the financial/controlling department in the establishment of appropriate provisions to cover these risks.

One of the major disadvantages of such systems is that they are usually quite complex. Understanding the whole process of how results are determined requires a deep technical and mathematical understanding and, most importantly, the necessary time. At this point, complexity may become a problem, as time usually is a scarce factor in management. Decision makers need to be able to trust the results, especially if important decisions are to be based on them. In an operative, real-life DSS, it is our experience that technology acceptance can be strengthened e.g. by a clearly arranged data quality assurance dashboard and, in general, reports whose visual appearance is not completely new to the receivers' eye. Subsequently, professional communication of project results, analyses and ultimately, between managers, analysts and developers is inevitable and plays an important role in earning the necessary confidence.

7 Conclusion

In the course of this paper, we investigated means to reduce the residual value risk car manufacturers and leasing companies face in their daily business. For this purpose, a DSS with forecasting capabilities was developed, implemented and tested. As accurate residual value forecasts are a critical success factor in this business, the focus has been put on suitable forecasting methodologies. ANNs have proven to be a reliable method for this purpose and thus, were implemented in the forecasting module of the DSS presented in this paper. It is necessary to mention that this system is in operative use and under constant improvement, as data is being automatically collected, cleansed, stored and used to improve the accuracy of future residual value forecasts. Furthermore, the flexibility of the integrated data warehouse and OLAP cube allow the investigation of a wide range of further, business-critical issues. From the performance measurement of dealerships to the impact of pricing strategies, promotions and discounts on residual values as well as interdependencies of the new and used car market, many future research opportunities are conceivable.

From a more general perspective, it can be concluded that corporations usually have access to large amounts of exclusive, unstructured, unintegrated data, which, properly processed, prepared and interpreted, have the potential to make a difference in daily business or even in important strategic decisions. This potential should not remain unexploited. Nevertheless, the information obtained still relies on humans with expert knowledge to interpret and communicate the results correctly. In a practical environment, close coordination between these experts and decision makers is crucial for the success of such projects. Since easy to understand reports can reduce complexity, build up trust and thus, facilitate communication, further research should also concentrate on proper visualization of data and analyses.

References

1. Leaseurope, 2014 leasing key facts & figures. http://www.leaseurope.org (Accessed: 19.07.2016)
2. DAT Group, DAT Report 2016. https://www.dat.de/report (Accessed: 16.07.2016)
3. Prado, S., Ananth, R.: Breaking through risk management, a derivative for the leasing industry. Journal of Financial Transformation, 34:211–218 (2012)
4. Nelson, M., Hill, T., Remus, W., O'Connor, M.: Time series forecasting using neural networks: should the data be deseasonalized first? J. of Forecasting, 18(5):359-367 (1999)

5. Lessmann, S., Listiani, M., Voß, S.: Decision Support in Car Leasing: A Forecasting Model for Residual Value Estimation. In: M. Lacity, F. Niederman, S. March (eds.) Proc. of the Intern. Conf. on Information Systems, Saint Louis, MO, USA (2010)
6. Wu, J. D., Hsu, C. C., Chen, H. C.: An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(4):7809–7817 (2009)
7. Lian, C., Zhao, D., Cheng, J.: A fuzzy logic based evolutionary neural network for automotive residual value forecast. In: Proceedings of the International Conference on Information Technology: Research and Education, pp. 545-548 (2003)
8. Prado, S.: The European used-car market at a glance: Hedonic resale price valuation in automotive leasing industry. Economics Bulletin, 29(3):2086- 2099 (2009)
9. Fan, H., AbouRizk, S., Kim, H., Zaïane, O.: Assessing residual value of heavy construction equipment using predictive data mining model. Journal of Computing in Civil Engineering, 22(3):181–191 (2008)
10. Rode, D. C., Fischbeck, P. S., Dean, S. R.: Residual risk and the valuation of leases under uncertainty and limited information. Journal of Structured Finance, 7(4):37–49 (2002)
11. Storchmann, K.: On the depreciation of automobiles: An international comparison. Transportation, 31(4):371–408 (2004)
12. Smith, L. D., Jin, B.: Modeling exposure to losses on automobile leases. Review of Quantitative Finance and Accounting, 29(3):241–266 (2007)
13. Cleveland, R. B., Cleveland, W. S., McRae, J. E., Terpenning, I.: STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1):3-73 (1990)
14. Broock, W. A., Scheinkman, J. A., Dechert, W. D., LeBaron, B.: A test for independence based on the correlation dimension. Econometric Reviews, 15(3):197-235 (1996)
15. Bishop, C. M.: Neural networks for pattern recognition. Oxford University Press (1995)
16. Sermpinis, G., Stasinakis, C., Dunis, C.: Stochastic and genetic neural network combinations in trading and hybrid time-varying leverage effects. Journal of International Financial Markets, Institutions and Money, 30:21–54 (2014)
17. Zimmermann, H., Neuneier, R., Grothmann, R.: Multi-agent modeling of multiple FX-markets by neural networks. IEEE Transactions on Neural Networks, 12(4):735–743 (2001)
18. von Spreckelsen, C., von Mettenheim, H.-J., Breitner, M. H.: Real-time pricing and hedging of options on currency futures with artificial neural networks. Journal of Forecasting, 33(6):419–432 (2014)
19. Eilers, D., Dunis, C. L., von Mettenheim, H.-J., Breitner, M. H.: Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning. Decision Support Systems, 64:100-108 (2014)
20. Kuo, R. J., Chen, C. H., Hwang, Y. C.: An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. Fuzzy Sets and Systems, 118(1):21–45 (2001)
21. Schocken, S., Ariav, G.: Neural networks for decision support: Problems and opportunities. Decision Support Systems, 11(5):393–414 (1994)
22. Zhang, G., Patuwo, B.E., Hu, M.Y.: Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14:35–62 (1998)
23. Adya, M., Collopy, F.: How effective are neural networks at forecasting and prediction? A review and evaluation. Journal of Forecasting, 17(5-6):481-495 (1998)
24. Mettenheim, H.-J., Breitner, M.: Robust decision support systems with matrix forecasts and shared layer perceptrons for finance and other applications. Proceedings of the International Conference on Information Systems, Saint Louis, MO, USA (2010)
25. Diebold, F. X., Mariano, R. S.: Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3):253–263 (1995)