International Journal of Forecasting 33 (2017) 864–877

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy

Stefan Lessmann [a,*], Stefan Voß [b]

[a] School of Business and Economics, Humboldt-University of Berlin, Unter-den-Linden 6, 10099 Berlin, Germany
[b] Institute of Information Systems, University of Hamburg, Von-Mell-Park 5, 20146 Hamburg, Germany

ARTICLE INFO

Keywords:
Forecasting
Data mining
Decision support systems
Automotive industry

ABSTRACT

The paper investigates statistical models for forecasting the resale prices of used cars. An empirical study is performed to explore the contributions of different degrees of freedom in the modeling process to the forecast accuracy. First, a comparative analysis of alternative prediction methods provides evidence that random forest regression is particularly effective for resale price forecasting. It is also shown that the use of linear regression, the prevailing method in previous work, should be avoided. Second, the empirical results demonstrate the presence of heterogeneity in resale price forecasting and identify methods that can automatically overcome its detrimental effect on the forecast accuracy. Finally, the study confirms that the sellers of used cars possess informational advantages over market research agencies, which enable them to forecast resale prices more accurately. This implies that sellers have an incentive to invest in in-house forecasting solutions, instead of basing their pricing decisions on externally generated residual value estimates.

1. Introduction

The paper explores the potential of forecasting methods to support decision making in the automotive industry. More specifically, we concentrate on the second-hand market and develop empirical models for forecasting resale prices. Given that the sale of new cars is typically associated with taking back used vehicles due to, e.g., retail trade-ins, repossessions and fleet returns from car rental companies (e.g., Du et al., 2009), the used car market is strategically important for car manufacturers.

Forecasting[1] is a popular approach to improving business processes and supporting decision making (e.g., Cang & Yu, 2014; Ha & Krishnan, 2012). In particular, we develop forecasting models for supporting pricing decisions. A considerable body of research has shown that sophisticated pricing strategies can increase the profitability of customer-centric operations substantially (e.g., Mantrala et al., 2006; Sharif Azadeh et al., 2015). For example, Marn et al. (2003) estimate that a 1% increase in sales prices can translate into an 8% increase in operational profits for an average S&P 500 company. Pricing is especially important in the used car market. Given that the quantity is largely fixed (i.e., because of take-back obligations), the price is the only control variable for increasing sales revenue and profit (Du et al., 2009).

In general, making effective pricing decisions requires a good estimate of the demand (e.g., Ferrer et al., 2012). In the second-hand car market, the demand depends substantially on the difference between a car's residual value and its offer price (Jerenz, 2008). Thus, when deciding on

* Corresponding author. Fax: +49 030 2093 5741.
E-mail addresses: stefan.lessmann@hu-berlin.de (S. Lessmann), stefan.voss@uni-hamburg.de (S. Voß).

[1] We do not distinguish between forecasting and predictive modeling here. Accordingly, the terms 'forecast' and 'prediction' are used interchangeably in the paper. Furthermore, we assume that forecasting models can be distinguished into regression and classification models that forecast continuous and discrete dependent variables, respectively. The paper is concerned with empirical methods of developing regression methods.

http://dx.doi.org/10.1016/j.ijforecast.2017.04.003
0169-2070/

offer prices, sellers need to estimate both residual values and resale prices. Such forecasts are also important in the new car business, where leasing has become a major sales channel (e.g., Pierce, 2012). Leasing companies set prices (i.e., leasing rates) based on expected residual values (e.g., Desai & Purohit, 1998). If the actual resale price of a car falls below expectations, the company faces a loss. Consequently, the accuracy of resale price forecasts is linked directly to the profitability of car leasing. More generally, given the impact of pricing policies on firms' performances and the dependence of pricing decisions on residual value estimates, we argue that resale price forecasting (RPF) is connected indirectly to the profitability of car selling. This suggests that attempts to increase the accuracy of resale price forecasts are managerially important and a relevant research topic.

Previous studies in the used car market have generally comprised a statistical modeling of resale prices but have rarely taken a decision support perspective. For example, only two studies have used a forecasting method other than multivariate linear regression. To the best of our knowledge, no determinants of the forecast accuracy other than the forecasting method have been explored at all (see Section 2 for details). The objective of the paper is to fill this gap. The corresponding results support car manufacturers who remarket used vehicles on a large scale, for example in conjunction with a leasing business, in their price management, as well as independent car dealers. More specifically, sellers need to decide how to obtain car residual values as inputs for their pricing strategy. One option is to purchase car residual value estimates from market research agencies such as ALG.[2] Alternatively, sellers can set up in-house forecasting support systems (FSS) for estimating prices. The advantages of the latter approach include full transparency and control over price estimation, and the opportunity to use all available information, including specific car characteristics. The paper clarifies the potential of such information to increase the accuracy of car resale price forecasts and, more generally, provides insights as to how best to approach the car resale price forecasting task.

The paper makes three contributions in pursuing its objectives. First, we perform a systematic comparison of several widely diverse forecasting methods under different experimental conditions. This is useful for identifying methods that are particularly well-suited for RPF and for augmenting the findings from other domains concerning the relative effectiveness of alternative forecasting methods. Second, we examine the predictive value of car-specific information and find that such (private) information allows used car dealers to increase the accuracy of their resale price forecasts. This demonstrates the merit of in-house forecasting support systems (FSS) compared to relying exclusively on purchased residual value estimates. Third, we examine the amount of effort required to update and manage forecasting models as part of an in-house solution. Such an analysis provides insights into the economic rationality of in-house FSS and helps managers to make informed decisions. In summary, our study delivers original findings concerning various methodological, organizational, and economic aspects of RPF.

The paper is organized as follows. We review related literature in Section 2 and develop our research questions in Section 3. We then describe our methodology in Section 4 and report experimental results in Section 5. Section 6 concludes the paper. The online appendix[3] provides additional results and describes the forecasting methods considered in the study.

2. Related work

We organize the related literature into two categories. From an application perspective, papers that consider the second-hand car market are related to this work. From a methodological point of view, our study is related to the forecasting literature.

The second-hand car market is a popular object of scientific enquiry. Many studies have focused on the informational efficiency of the market; that is, how efficiency changes over time and especially in response to online-mediated sales channels (e.g., Adams et al., 2011; Andrews & Benzing, 2007; Genesove, 1993). Market prices play a central role in such work, as they determine the informational efficiency of the market and are a key element in consumer decisions.

Market prices are also relevant for decision support. In particular, price forecasts are a crucial input for revenue management initiatives and advanced selling support systems (Sharif Azadeh et al., 2015). However, only a few studies have emphasized revenue management, and more specifically the connection between pricing decisions and revenue, in the used car business. Jerenz (2008) develops a pricing optimization system that consists of three parts: a forecasting component for estimating residual values, a statistical model for estimating price response functions (given model-estimated residual values), and a dynamic optimization engine for determining the optimal pricing strategy. Du et al. (2009) propose a related three-stage approach for identifying the optimal distribution of auction vehicles. In their approach, the expected auction price of a used vehicle and the local market elasticity are modeled using linear regression and an ARIMA process, respectively. The corresponding results represent the input to a genetic algorithm, which maximizes the net auction profit of distributing vehicles on the basis of their estimated auction prices, asset carrying costs and business constraints (Du et al., 2009).

Both studies illustrate that resale price forecasts form a starting point for systematic revenue management initiatives. More specifically, they exemplify how the effectiveness of a used car selling support system depends on the accuracy of the resale price forecasts. This suggests that better (more accurate) forecasts could make a sizeable contribution to the effectiveness of such system. However, Du et al. (2009) and Jerenz (2008) both model resale prices by means of multivariate linear regressions. Although *explanatory* regression models possess many advantages, they forecast less accurately than data-driven

[2] http://www.alg.com.

[3] Available as an online supplement (see Appendix A).