European Journal of Operational Research 288 (2021) 466–479

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Production, Manufacturing, Transportation and Logistics

Using shared sell-through data to forecast wholesaler demand in multi-echelon supply chains

Jente Van Belle*, Tias Guns, Wouter Verbeke

Solvay Business School, Data Analytics Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, Brussels 1050, Belgium

ARTICLE INFO

Article history:
Received 19 May 2019
Accepted 29 May 2020
Available online 5 June 2020

Keywords:
Forecasting
Bullwhip effect
Information sharing
Sell-through data
Machine Learning

ABSTRACT

Operational forecasting in supply chain management supports a variety of short-term planning decisions, such as production scheduling and inventory management. In this respect, improving short-term forecast accuracy is a way to build a more agile supply chain for manufacturing companies. Demand forecasting often relies on well-established univariate forecasting methods to extrapolate historical demand. Collaboration across the supply chain, including information sharing, is suggested in the literature to improve upon the forecast accuracy of such traditional methods. In this paper, we review empirical studies considering the use of downstream information in demand forecasting and investigate different modeling approaches and forecasting methods to incorporate such data. Where empirical findings on information sharing mainly focus on point-of-sale data in two-level supply chains, this research empirically investigates the added value of using sell-through data originating from intermediaries, next to historical demand figures, in a multi-echelon supply chain. In a case study concerning a US drug manufacturer, we evaluate different methods to incorporate this data and consider both time series methods and machine learning techniques to produce multi-step ahead weekly forecasts. The results show that the manufacturer can effectively improve its short-term forecast accuracy by integrating sell-through data into the forecasting process and provide useful insights as to the different modeling approaches used. The conclusion holds for all forecast horizons considered, though it is most pronounced for one-step ahead forecasts. Therefore, our research provides a clear incentive for manufacturers to assess the forecast accuracy that can be achieved by using sell-through data.

1. Introduction

Supply chain planning is defined as “the forward-looking process of coordinating assets to optimize the delivery of goods, services and information from supplier to customer, balancing supply and demand” (Gartner Inc., 2019). The driving, or rather pulling, force of the supply chain is the final customer demand. In order to fulfill this customer demand, generally, orders are placed with the next level upstream[1] in the supply chain. Depending on the length of the supply chain, this demand, in turn, generates orders at the next level(s). Fulfillment of these orders generates a downstream flow of goods which complements the upstream flow of demand information. Fig. 1 depicts a multi-echelon supply chain, with goods and information flows represented by solid and dashed lines, respectively. In this multi-echelon supply chain setting, customer demand, represented by the point-of-sale (POS) information flow, will deplete the retailer’s inventory. As the retailer wants to maintain a desired level of inventory, it will place an order with a wholesaler to replenish its stock, which in turn will (eventually) result in the wholesaler placing an order with the manufacturer. Note that there are typically many wholesalers that serve many retailers. In this setup, product-related information that is directly available to the wholesaler is referred to as sell-through data from the manufacturer’s perspective. This naturally includes data on retailer demand but also wholesaler inventory positions, among others. From the manufacturer’s perspective, sell-through and POS data are two different types of downstream information.

Effective supply chain planning is of great importance to manufacturing firms on all planning levels. In this paper we focus on the operational level of supply chain planning which encompasses a variety of short-term decisions, such as raw materials procurement, inventory management and production scheduling. Operating in an increasingly competitive environment, today, manufacturing companies are organized following the principles

* Corresponding author.
E-mail addresses: Jente.Van.Belle@vub.be (J. Van Belle), Tias.Guns@vub.be (T. Guns), Wouter.Verbeke@vub.be (W. Verbeke).

[1] We opt to characterize the supply chain in this particular vertical order, where the final customer is the most downstream element.

https://doi.org/10.1016/j.ejor.2020.05.059

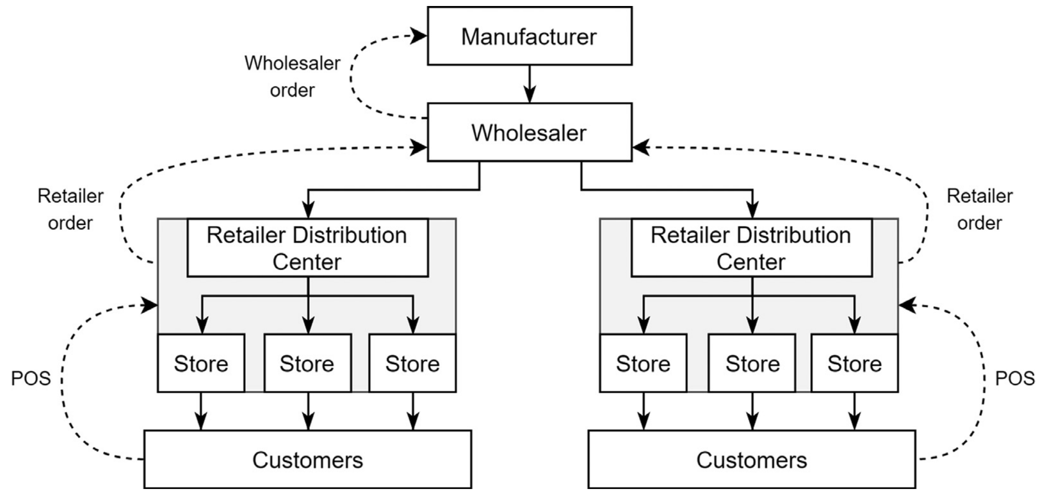

Fig. 1. Flows of goods (—–), and information (- - -) in a multi-echelon supply chain.

of just-in-time (JIT) delivery which focuses on efficiency and eliminating the waste of excessive inventory. In JIT systems, accurate Stock Keeping Unit (SKU) short-term demand forecasts, typically on a daily, weekly or monthly basis, are crucial as they are the fundamental inputs to drive the supply chain planning system. Therefore, forecasting performance can strongly affect a company's financial results, since both under- and overestimating short-term demand are expensive as they result in supply shortages and consequently poor customer service, and excess inventories and product obsolescence, respectively (Sanders and Graman, 2009).

Traditionally, demand forecasting relies on univariate time series methods that analyze prior demand data[2] in order to extrapolate the extracted demand pattern to the future. Note that prior demand data here solely refers to the demand directly observed by the supply chain participant producing the forecasts. In the multi-echelon supply chain setup, depicted in Fig. 1, this implies that manufacturers use information on incoming wholesaler demand. It is well known that this demand information is a distorted version of customer demand as demand information becomes increasingly altered and volatile moving upstream in the supply chain. This effect is known as the bullwhip effect (BWE) and its study can be traced back to Forrester (1961). Using this volatile demand information can lead to poor forecasts and subsequently supply chain inefficiencies. In their seminal papers, Lee, Padmanabhan, and Whang (1997a, 1997b) identify four major causes of the BWE: demand signal processing, order batching, rationing and shortage gaming, and price fluctuations and promotions. For empirical results on the measurement of the BWE; see e.g., Fransoo and Wouters (2000).

The aim of this work is to tackle the first cause of the BWE, i.e., demand signal processing. Recall that each supply chain participant usually takes only directly observed demand into account in forecasting demand. If the supply chain participants are forward-looking[3], at the end of each period, they observe their most recent demand and update their forecasts and target inventory levels accordingly. By regularly updating forecasts and target inventory levels, the variability in demand observed by the next level upstream increases. This effect is propagated and amplified through the chain as this higher-variability demand information is used as input to the demand forecasting process at this next level upstream. Demand signal processing in this traditional manner also implies that there is an information lead time: the more upstream the supply chain participant is, the slower the changes in final customer demand will appear in its observed demand. Order batching, rationing and shortage gaming, and price fluctuations and promotions can all further magnify the increase in variability.

Previous work on the BWE has led to a number of suggestions for reducing its impact. One possible solution proposed by Lee et al. (1997a) is to share information on an interorganizational level: "one remedy [...] is to make demand data at a downstream site available to the upstream site." As pointed out by Byrne (2012), upstream supply chain participants can sense final customer demand volatility more quickly in this manner, provided that the downstream demand data is closely related to final customer demand. By reducing uncertainty and removing delays in translating the demand signal, an improvement in forecast accuracy can be expected. However, there has been an ongoing debate in the literature whether information sharing is necessary to counter the impact of the BWE. Whilst the discussion is mainly focused around analytical results that are built on restrictive assumptions, the limited number of data-driven empirical studies mostly provide evidence on the benefits of using downstream demand information in demand forecasting. An overview of these empirical studies is provided in Section 2.2. Finally, note that other types of information can be used as well to improve forecast accuracy and mitigate the impact of the BWE. See e.g., Trapero, Pedregal, Fildes, and Kourentzes (2013), Trapero, Kourentzes, and Fildes (2015) and Kourentzes and Petropoulos (2016) for studies on the use of promotional data.

The empirical studies, discussed in Section 2.2, focus on using POS data in two-level supply chains. In multi-echelon supply chains, as depicted in Fig. 1, we can also use the downstream information originating from intermediaries: sell-through data. In theory, we could use both sources of downstream demand data, but by using POS data we should be able to capture leading dynamics sooner than when we use sell-through data, provided that the POS data well approximates final customer demand. In reality, however, it is often difficult to obtain accurate POS data for short-term forecasting purposes in time. For instance, in the pharmaceutical industry, POS level information is often only available with a delay via a third-party commercial provider of market intelligence data. On the other hand, sell-through data is often directly available via

[2] Typically sales figures are being used as an approximation to demand (and therefore the terms demand and sales are often used interchangeably), as actual demand is unknown because of, inter alia, stock-outs. However, also orders or shipments can be used to approximate demand.

[3] If the supply chain participants are not forward-looking and thus do not update their demand forecasts and target inventory levels in each period, demand signal processing would not add to the BWE.

Electronic Data Interchange (EDI) transactions. Lee et al. (1997a) argue that one can then resort to sell-through data in order to mitigate the impact of the BWE. This is the setting that we consider in this paper. However, note that there are also cases in which POS data may be available, but sell-through data is potentially more useful to the manufacturer. Two important examples are an SKU for which retailers experience frequent stockouts and a substitute product is available, and a heavily promoted SKU for which retailers do not share information on promotional activities with upstream supply chain participants. In the first example, POS data is obviously not a good approximation of final customer demand, while sell-through data may possibly contain relevant information to the manufacturer if retailers respond to stockouts by placing larger orders with the wholesaler. In the second example, POS data contains lagging instead of leading dynamics; however, as retailers place larger orders with wholesalers in anticipation of promotional activities, downstream demand information that is included in sell-through data implicitly captures these promotional effects.

Our contributions in this paper are as follows:

- We present a broad review of the empirical literature on using downstream information in demand forecasting with a focus on the different modeling approaches used and the forecast horizons considered.
- We evaluate the use of sell-through data as an alternative to POS data, and investigate different modeling approaches and forecasting methods to incorporate such data. In this respect, where previous studies solely focused on downstream demand as such, we also consider information on downstream inventory positions.
- We empirically analyze a range of methods that allow us to incorporate sell-through data, including time series methods and machine learning techniques, and compare them with traditional univariate forecasting methods. To this end, we use data from a US drug manufacturer for 50 different SKUs and address both one-step and multi-step ahead forecasts.
- We highlight differences between the different modeling approaches and discuss their performance in relation to forecast horizon and the associated implications for the availability of downstream information.

This paper is structured as follows. In the next section, we review relevant literature. In Section 3, we survey baseline forecasting methods and methods that allow us to include exogenous data sources such as downstream information. The case study and experimental design are introduced in Section 4. Results are presented and discussed in Section 5 with concluding remarks following in Section 6.

2. Using downstream information in demand forecasting

In this section, we first define the scope of our study within the literature on information sharing and briefly address the ongoing discussion based on theoretical, conceptual and empirical findings. Next, we survey and review data-driven empirical studies upon which our work expands.

2.1. Information sharing

Owing to the suggestion of, among others, Lee et al. (1997a), to share information on an interorganizational level to deal with the impact of the BWE, various degrees of information sharing have been explored in the literature. Holweg, Disney, Holmström, and Småros (2005) identify four different types of supply chain collaboration: (i) the traditional supply chain, where there is no formal collaboration, i.e., no sharing of information; (ii) information exchange, where there is exchange of demand (and other product-related) information; (iii) vendor managed replenishment, where the vendor takes responsibility for maintaining the client's inventory; and (iv) synchronized supply, where the supplier manages the whole chain on the operational level directly. This research focuses on the first and second type of collaboration. However, note that research on information sharing in a supply chain context is not limited to the use of downstream information to improve demand forecasting. See e.g., the work of Zhang, Tan, Robb, and Zheng (2006) on the value of sharing shipment information, where one stage in a supply chain shares shipment quantity information with its immediate downstream customers, a practice also known as advanced shipping notice.

The academic literature on the use of information exchange to deal with the impact of the BWE is ambiguous. This ongoing debate mainly exists of a vast amount of theoretical studies based on analytical models and simulated data. However, in order for these models to be tractable, they tend to be highly modified versions of reality as they are built on restrictive assumptions (e.g., generally they do not include the effects of price fluctuations and promotions, although these are identified as possible causes of the BWE). Some theoretical studies have shown downstream information to be valuable to the supplier. In this regard, Chen, Drezner, Ryan, and Simchi-Levi (2000) show that, under certain assumptions, and if each stage of the supply chain has complete knowledge of the final customer demand, information exchange will reduce the impact of the BWE although it will not completely eliminate it. On the other hand, several studies indicate that through exploitation of an appropriate forecasting model, in certain specific conditions, it may be possible to infer all necessary information about the downstream demand without using this information directly, thus suggesting demand information exchange to become redundant; see e.g., Raghunathan (2001). We refer to Trapero, Kourentzes, and Fildes (2012) and Syntetos, Babai, Boylan, Kolassa, and Nikolopoulos (2016) and the references therein for elaborate discussions on this subject.

A second stream of literature addresses the BWE using conceptual modeling. Lee, Kim, and Kim (2014) develop a conceptual framework of interorganizational systems visibility, which refers to the extent to which information required for effective supply chain cooperation is shared between partner firms through interorganizational information systems. By testing their conceptual framework, they conclude that there is a significant positive relationship between supply chain performance and the degree of interorganizational systems visibility. In the same spirit, Kim, Ryoo, and Jung (2011) show that the buyers, in particular, need to make their internal information systems visible to suppliers, so that they can have timely access to the integrated information for informed decision making and as such improve supply chain performance.

Finally, a third stream of literature involves data-driven empirical research. These studies, upon which our work expands and which are reviewed in detail in the next section, mostly provide evidence on the benefits of using downstream information in demand forecasting.

2.2. Empirical research

In Table 1, we provide an exhaustive (to the best of our knowledge) chronological overview of empirical studies concerning the use of downstream information in demand forecasting. We divide these studies into two categories, namely the substitution and integration approach, depending on the modeling approach that is applied. Studies belonging to the first category capture information exchange by substituting the directly observed prior demand by the downstream demand information. This approach can be applied if all goods pass through intermediaries, i.e., if there are no direct sales from the manufacturer to final customers. In such sup-

Table 1
Literature review on demand forecasting with information sharing.

Author(s), Year	Target context	Nr. of targets	Target frequency[a]	Sample size	Downstream information type	Forecast horizons	Modeling approach[b]	Category[c]
Hanssens (1998)	High-Tech	1	M/Q	44/11	POS	1	ECM	2
Byrne and Heavey (2006)	Industrial	16	D	260	POS	1	Simulation & TS	1
Hosoda et al. (2008)	Retail	3	W	51	POS	1	TS	1
Kelepouris et al. (2008)	Retail	48	W	52	POS	1	Simulation & TS	1
Williams and Waller (2010)	Retail	432	W	104	POS	4, 13 & 26	TS	1
Williams and Waller (2011)	Retail	10 & 180	W	104	POS	1–13	TS & hierarchical	1
Trapero et al. (2012)	Retail	43	W	52	POS	1	TS, TSX & NN	2
Williams et al. (2014)	Retail	36	W	110	POS	1–6	TS & ECM	1 & 2
Hartzel and Wood (2017)	Retail	12.350	W	36	POS	1	NN	2
Our study	Pharma	50	W	126 – 253	*Sell-through*	1–5	TS, TSX & ML	1 & 2

[a] D = daily, W = weekly, M = monthly, Q = quarterly.
[b] ECM = error correction modeling, TS = extrapolative time series methods, TSX = extrapolative time series methods with exogenous variables, NN = neural networks, ML = machine learning techniques.
[c] Cat. 1 = substitution approach, Cat. 2 = integration approach.

ply chains, use of this substitution approach is justified as both demand signals, the directly observed and downstream demand, capture all flows of goods and therefore result in equal aggregates over time. The substitution approach relies on the idea that the downstream demand signal may be a better predictor of future directly observed demand as it is generally less variable than directly observed prior demand. Studies belonging to the second category use forecasting methods that integrate both sources of information, i.e., the traditional prior demand and downstream information. We additionally provide information on the context and the number of target time series considered. Information regarding frequency, sample size, type of downstream information used, forecast horizon and modeling approach are provided as well. Finally, we highlight in Table 1 how our work extends the existing studies.

Category I: Substitution approach

Research concerning information exchange mainly involves analytical studies characterized by restrictive assumptions. An early study that expands upon these restrictive analytical studies to incorporate real world complexities into the analysis is that of Byrne and Heavey (2006). They model and analyze the effect of information exchange and forecasting on the performance parameters of an actual industrial supply chain involving multiple customers, distributors and product families. Through a simulation study, they highlight the significant benefits that are achievable through the use of improved information exchange and forecasting methods. Potential total supply chain cost savings are shown to amount to 9.7%.

Hosoda, Naim, Disney, and Potter (2008) emphasize the discrepancy between theoretical models and empirical findings. They investigate the benefit of sharing POS information in a two-level supply chain based on real demand data, where a theoretical model argues there is no benefit from such a collaboration scheme. They show that significant benefits can be achieved by exploiting POS data. In a similar setting, also Kelepouris, Miliotis, and Pramatari (2008) find that information sharing enabled the upstream supply chain member to respond more effectively and efficiently to demand variations, to produce more accurate forecasts, and therefore to reduce the impact of the BWE and inventory levels.

Williams and Waller (2010) come to similar conclusions but emphasize that, although the use of POS data as forecasting input produces a lower forecast error on average and in the majority of the cases, it does not always outperform the use of traditional data in terms of forecast accuracy. Their results are based on several SKUs in a retail business environment. In a similar setting, Williams and Waller (2011) explore the interplay between the use of POS data and different hierarchical forecasting approaches.

Category II: Integration approach

While the transition from category I to category II studies evolved in time, the earliest empirical study listed in Table 1 integrates both sources of information. Hanssens (1998) studies the use of POS data in the context of a durable product in the category of personal computing accessories. The channel structure for this product has either one or two intermediaries between manufacturer and end-user. However, information on intermediate levels in the channel is not available. Based on an error-correction model, which takes into account both prior demand and POS data, statistical evidence shows that the use of POS information significantly improves the accuracy of the demand forecasts at manufacturer level. An improvement of 63% in out-of-sample accuracy is obtained over a univariate baseline. Trapero et al. (2012) show as well that suppliers can achieve substantial improvements in forecast accuracy by using POS information. They use a multivariate autoregressive model with exogenous inputs and neural networks to integrate both sources of information in a serially linked two-level supply chain.

Building on the work of Hanssens (1998), Williams, Waller, Ahire, and Ferrier (2014) present evidence that retail echelon inventory processes translate into a long-run equilibrium between orders, which are used to approximate prior demand, and POS data. If this long-run equilibrium did not exist, the retailer's inventory would grow without bound, or backorders and lost sales would accumulate to unacceptable levels. They refer to this long-run equilibrium as the inventory balance effect. Simultaneously using both sources of information and their interrelation is shown to outperform approaches based only on prior demand or POS data in predicting retailer orders to suppliers for consumer goods. This study thus relies on forecasting methods belonging to both the integration and substitution approach. However, for the substitution approach, only the additive Holt-Winters exponential smoothing method is considered.

Like Trapero et al. (2012), also Hartzel and Wood (2017) adopt neural networks for forecasting. They achieve an overall 11.2% improvement over forecasting using only the prior demand data. Moreover, they identify several factors that affect the degree of improvement by using POS data: item order quantity, item order variability and item order frequency[4].

All previous studies use POS data as downstream data source. However, it is not always possible to obtain accurate POS data for short-term forecasting purposes in time. As indicated by Hanssens (1998), in the retail business, accurate customer demand data is typically readily available through consolidated scanning services

[4] They use prior orders to approximate demand.

and instant demand feedback. However, in many industries, like the pharmaceutical sector, this is not the standard (yet), which makes it hard to gather accurate customer demand data. In multi-echelon supply chains, manufacturers can also use sell-through data in search for improved forecast accuracy. In contrast to POS data, this sell-through data is often directly available via EDI transactions.

Longer forecast horizons

Most studies only focus on the impact of using POS data on one-step ahead predictions. In practical applications, however, also longer horizons within the short-term are usually relevant as different horizons have different implications for the supplier in terms of the ability to improve operational decisions. In the overview presented in Table 1, only three studies consider forecast horizons $h > 1$, and only two of them allow the assessment of the effect of different horizons.

In the study of Williams and Waller (2010), mixed results are obtained with regard to the impact of the forecast horizon. They use a 4, 13 and 26 week forecast horizon and find strong statistical evidence for the shortest horizon. However, they also find that as the forecast horizon increases, the POS approach is the best in more cases. Taking into account both observations, they suggest that the improvements in accuracy by using the POS approach will be rather limited for longer horizons, which can be explained by the fact that both signals (POS and prior demand) are more closely aligned in the long run. As discussed in the previous section, this relation is formalized by Williams et al. (2014). They report that both their inventory balance approach and the POS approach lower the forecast error relative to the prior demand approaches over each week in the short-term forecast horizon from $h = 1$ to 6 and that their inventory balance approach consistently outperforms the POS approach. However, for both methods the relative gain decreases as the forecast horizon increases.

Research gap

The empirical literature on the use of downstream information in demand forecasting mainly focuses on case studies in the retail business, where POS data is often readily available. Also, most of these existing studies either consider information sharing by substituting the input or by integrating both sources of demand information. In this study, we consider both approaches and elaborate a case study in the pharmaceutical industry, where POS data is usually not readily available but manufacturers can resort to sell-through data in search for improved forecast accuracy. For the integration approach, we rely on time series methods and machine learning techniques, and where previous studies solely focused on downstream demand as such, we also consider information on downstream inventory positions. Finally, because of its practical relevance, we consider one- to five-step ahead weekly forecasts.

3. Review of forecasting methods

The categorization of empirical studies on the use of downstream information in demand forecasting (see Table 1) shows that various modeling approaches have been evaluated. In this regard, quantitative demand forecasting methods can roughly be divided in two groups: extrapolative time series and explanatory methods. Methods belonging to the first group only use information on the variable to be forecast. The studies, listed in Table 1, in which the substitution approach is adopted rely on extrapolative time series methods. Explanatory methods, on the other hand, explicitly take into account the relationships between the variable to be forecast and one or more other variables. This approach is used in the studies that integrate both sources of information to include the exogenous downstream information in forecasting demand. In this section, we first briefly review well-known extrapolative time series methods. Second, we discuss explanatory methods and distinguish between (i) extensions to extrapolative time series methods developed to allow for exogenous variables, and (ii) machine learning techniques.

3.1. Extrapolative time series methods

Traditional forecasting methods are based on modeling the past time series structure and extrapolating it into the future. The study of these univariate time series methods is extensive with the exponential smoothing (ETS) and ARIMA methods being the most well-known (Ord, Fildes, and Kourentzes, 2017). Automatic model selection algorithms based on minimizing some information criterion are available for both ARIMA and ETS (Hyndman and Khandakar, 2008; Svetunkov, 2020). In the recent M4 Competition (Makridakis, Spiliotis, and Assimakopoulos, 2020), ETS and ARIMA served as 'standards for comparison' because of their widespread use in practice and showed relatively good forecast accuracy.

3.2. Explanatory methods

Extrapolative time series methods with exogenous variables

Prime candidate approaches for explanatory modeling in a time series context are the ETS and ARIMA methods extended to include exogenous variables. In both cases, this can be achieved in two ways. A first approach relies on enhancing the ETS and ARIMA frameworks to include additive exogenous effects directly; see e.g., Hyndman, Koehler, Ord, and Snyder (2008), Athanasopoulos and Hyndman (2008), Trapero et al. (2013), and Kourentzes and Petropoulos (2016) for ETS and Trapero et al. (2012) and Box, Jenkins, Reinsel, and Ljung (2015) for ARIMA. A second approach is to construct a regression model and use ARIMA or ETS to model the error term of the regression model; see e.g., Trapero et al. (2015) for regression modeling with ARIMA errors. Note that it is also possible to reverse this order and use the exogenous variables to model any unexplained variation left in the residuals of the time series model.

In a time series context, exogenous variables may not immediately impact the time series to be forecast. Therefore, often time shifts (lags) of the exogenous variables are considered to model any leading dynamics to the target variable. If useful exogenous variables (and their relevant lag orders) can easily be identified, the ETS and ARIMA methods extended to include exogenous variables can readily be used to include exogenous information in the forecasting process. Often, however, it is unknown in advance whether potential exogenous variables do have predictive power nor with which lag they affect the time series of interest. As the aforementioned methods are essentially specific types of regression, including too many exogenous variables may result in overfitted models and poor forecasts. Therefore, it is paramount to identify and select the most relevant and complementary set of exogenous variables.

Machine learning techniques

As an alternative to the methods discussed in the previous section, we can also use machine learning (ML) techniques to include exogenous information in the forecasting process. Applying ML techniques to time series problems generally requires the construction of features characterizing the time series under consideration, including time shifts (lags) of the series itself. In this regard, ML techniques can be used for univariate time series modeling as well. However, the results of Makridakis, Spiliotis, and Assimakopoulos (2018) and the M4 competition (Makridakis et al., 2020) suggest that statistical time series methods generally outperform ML techniques in a univariate setting. Nevertheless, the usefulness of ML techniques in an explanatory forecasting setting results from the fact that an arbitrary number of input features can be specified. For identifying the most relevant input variables and their respective lag orders, these ML techniques often rely on the principle of regularization to deal with this problem directly.

Several ML techniques have been applied in a forecasting context with exogenous variables in the literature, ranging from regularized regression techniques to neural networks. Regularized regression is often implemented via the LASSO (Hastie, Tibshirani, & Friedman, 2011). LASSO has been successfully applied in an operational and tactical demand forecasting context with exogenous information by e.g., Huang, Fildes, and Soopramanien (2014), Ma, Fildes, and Huang (2016) and Sagaert, Aghezzaf, Kourentzes, and Desmet (2018). Other commonly used ML techniques are the multilayer perceptron neural network (Zhang, Patuwo, & Hu, 1998), support vector regression (Smola & Schölkopf, 2004) and (ensemble) tree methods (Breiman, 2001; Breiman, Friedman, Stone, & Olshen, 1984). The above techniques are applied to demand forecasting problems with exogenous information by e.g., Ali, Sayın, Van Woensel, and Fransoo (2009), Trapero et al. (2012) and Di Pillo, Latorre, Lucidi, and Procacci (2016).

4. Empirical study

In this section, we empirically evaluate the value of incorporating sell-through data in demand forecasting at manufacturer level. To this end, we collected data from a US drug manufacturer. In line with the survey of empirical studies covered in Section 2.2 and in order to quantify potential benefits, we compare the case when no information is shared and the manufacturer only has access to the incoming wholesaler demand with both the substitution and integration approach for the information sharing case. As discussed above, for the latter approach, we use both extrapolative time series methods with exogenous variables and ML techniques.

In this section, we first provide information on the dataset under study. Next, we outline the experimental setup by providing further details on the modeling approaches and the evaluation scheme used in the empirical evaluation.

4.1. Data

The involved drug manufacturer operates in a multi-echelon supply chain as depicted in Fig. 1, with all goods passing through wholesalers and retailers before reaching the final customers. As we focus on the use of sell-through data, the manufacturer can use two sources of information: (i) the prior shipments to wholesalers and (ii) sell-through data. The first source of information is collected by the information system of the manufacturer, while the latter is collected by the wholesalers and provided to the manufacturer via Electronic Data Interchange (EDI) transactions. The data have been collected on a weekly basis between January 2014 and October 2018. Over this 253-week period, a total of 50 SKUs were observed. As sales of some products were discontinued during this period, while other products were launched after January 2014, the SKUs have a variable number of observations. On average, we have 205 observations per SKU, with a minimum of 126. Also, for an average SKU, periods in which no manufacturer shipments are observed only account for 3.7% of the observations, with a maximum of 25.4%.

For each SKU we have four time series available. The first one corresponds to the prior shipments at manufacturer level, used as a proxy for wholesaler demand, while the other three are part of the sell-through data. The sell-through data is aggregated over all wholesalers and includes (i) weekly sales figures, used as a proxy for retailer demand, (ii) ending inventory positions, and (iii) incoming quantity shortages for the reporting periods. Each data

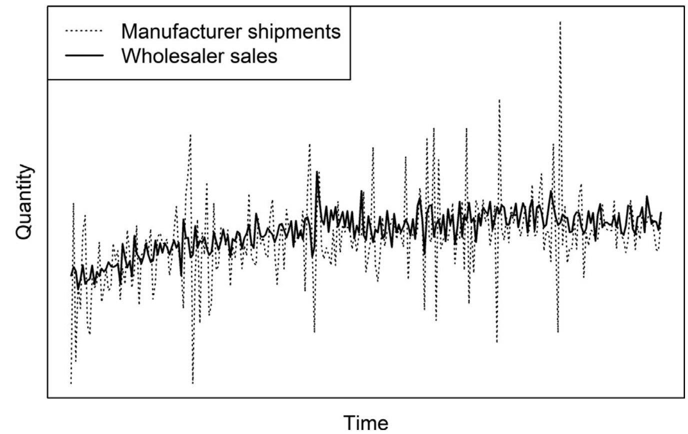

Fig. 2. Example SKU demand time series.

point in this last time series represents the total quantity that wholesalers expected to receive for the given reporting period but that was not yet received. Without information sharing, it is obvious that the manufacturer has no access to the wholesaler sales figures and ending inventory positions. In this case study, the same goes for the information on incoming quantity shortages as distribution and transportation to wholesalers is outsourced.

Recall that the motivation to use downstream information in demand forecasting is to mitigate the impact of the BWE. In Fig. 2, which shows an example SKU demand time series, we clearly observe the presence of the BWE[5]. A more formalized way of detecting the BWE was proposed by Fransoo and Wouters (2000). They measure the BWE by calculating the ratio of the coefficients of variation of the prior observed and downstream demand. In our specific case, this means that we can calculate the bullwhip ratio (BWR) as follows:

$$\mathrm{BWR} = \frac{\sigma_M/\mu_M}{\sigma_W/\mu_W} \tag{1}$$

where σ_i is the standard deviation and μ_i the mean, with $i = M$ representing manufacturer shipments and $i = W$ wholesaler sales, respectively. A BWR greater than one means that there is variance amplification, and thus provides evidence of the presence of the BWE. Fig. 3 shows a histogram of the BWR for our dataset and clearly indicates that there is variance amplification for all SKUs, though in varying degrees.

4.2. Modeling

In this section, we first discuss the methods that are used for the case when no information is shared. Next, we provide details on the methods used to incorporate the sell-through data, where we distinguish between the substitution and integration approach. Note that we do not make any assumptions regarding the data generating process of the observed demand patterns, not for the directly observed demand, nor for the information contained in the sell-through data. Therefore, a wide variety of forecasting methods can be adopted. An overview of the forecasting methods considered in this study, along with their respective inputs, is provided in Table 2.

4.2.1. No information sharing

The first baseline that is included is the Naïve method. This method simply takes the last known data point as the forecast

[5] In order to visualize the BWE, downstream incoming quantity shortages and ending inventory positions are not shown (as they have different ranges).

Table 2
Overview of forecasting methods and inputs.

Class	Method	Time series: Manufacturer shipments (y)	Time series: Wholesaler sales (w)	Time series features: AR terms (y) & trend	Time series features: Seasonal dummies	Sell-through information features
No information sharing (NIS)	Naïve	✓				
	ETS	✓			✓	
	ARIMA	✓			✓	
Information sharing (IS) Substitution	ETS-W		✓		✓	
	ARIMA-W		✓		✓	
Information sharing (IS) Integration - TS	ETSX	✓			✓	✓
	ARIMAX	✓			✓	✓
Information sharing (IS) Integration - ML	LASSO			✓	✓	✓
	MLP			✓	✓	✓
	SVR			✓	✓	✓
	RF			✓	✓	✓

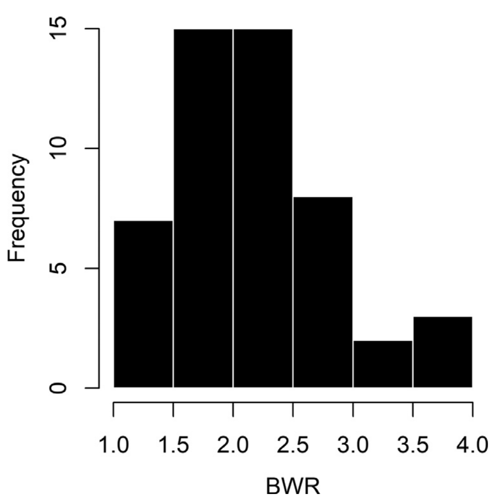

Fig. 3. Histogram of the BWR.

for the next periods and as such does not require parameter estimation. Therefore, a more complex method should outperform the Naïve one in order to justify the additional complexity.

As the aim is to assess whether sell-through data can enrich the forecasting process, also other (known) sources of variability should be captured by both the baseline(s) and the methods used for the case when information is shared. In business forecasting, a known source of variability is seasonality. Not incorporating seasonal information would possibly overestimate the predictive power of the sell-through data.

In this case study, we are dealing with weekly data over a relatively short time span (a maximum of 253 weeks). Therefore, we opt to not consider an annual seasonal pattern but only model month of year effects through the use of monthly dummy variables. In order to assign weeks (which are based on ISO 8601) to months, we use 13-week quarters and assign a month either 4 or 5 weeks, based on the number of weekdays in year 2017. In 2015, there were 53 weeks and we assign the extra week to the month December. Before adding the seasonal information to the forecasting models, we first make a selection of the seasonal variables via a hybrid stepwise selection strategy (Hastie et al., 2011) where we minimize the Akaike Information Criterion corrected for sample size (AICc) (Sugiura, 1978). More specifically, we start by regressing manufacturer shipments on an intercept only and then sequentially add to the linear model the variable that most improves the fit in terms of AICc. In each step, we also evaluate whether dropping a variable has a favorable effect on AICc. This stepwise search is terminated if no further improvement in AICc is possible. As potentially useful variables we consider the seasonal dummies and a linear trend variable (however, if the trend variable is selected, it is removed from the selection as trend is modeled differently by the different forecasting methods considered).

The extended versions of the time series methods, discussed in Section 3.2, allow us to incorporate the selected seasonal dummy variables. Therefore, as a second and third baseline, we opt for ETS and ARIMA. For these methods we rely on the automatic model selection algorithms implemented in the ‘smooth’ R package (Svetunkov, 2020). We only consider non-seasonal models as seasonal effects are modeled via the selected dummy variables and we do not allow for multiplicative trend in ETS models.

To predict future demand for the case when no information is shared we thus rely on prior shipments to wholesalers (y) and the set of selected seasonal dummy variables D^*:

$$\hat{y}_{t+h|t} = f(y_t, y_{t-1}, \ldots, y_{t-m}, D^*), \tag{2}$$

with m the number of observations in the in-sample period. However, note that D^* is omitted in the Naïve model.

4.2.2. Information sharing

Most existing empirical studies on information sharing, listed in Table 1, either consider information sharing by substituting the usual input, i.e., prior observed demand data, by downstream demand information in a time series model or by explanatory methods integrating both sources of demand information. In this study, both approaches are evaluated.

Substitution approach

For this first category of information sharing methods, we consider ETS and ARIMA where the downstream wholesaler sales time series (w) instead of the prior manufacturer shipments (y) is used as input:

$$\hat{y}_{t+h|t} = f(w_t, w_{t-1}, \ldots, w_{t-m}, D^*). \tag{3}$$

We name these models ETS-W and ARIMA-W. The set of seasonal dummy variables D^* is obtained by using the selection procedure described in Section 4.2.1 with manufacturer shipments (y) replaced by wholesaler sales (w). Recall that this substitution approach is only suited if all goods pass through wholesalers.

Integration approach

Within the second category of information sharing methods, we distinguish between extrapolative time series methods with exogenous variables and ML techniques. We consider two time series methods that allow modeling of exogenous variables and four ML techniques.

Extrapolative time series methods with exogenous variables – We consider both an ETS and ARIMA model with prior manufacturer shipments as input and with seasonal and sell-through information modeled as exogenous variables. To capture the downstream dynamics, we incorporate lagged versions of the three time series included in the sell-through data. Taking into account our goal of short-term forecasting and the perceived degree of operational responsiveness of the wholesalers in our case study, we consider lags up to order 5 for the wholesaler sales and ending inventory position. For the incoming quantity shortages, we only consider one lag. Recall that the incoming quantity shortage for a given reporting period represents the total quantity that wholesalers expected to receive for the reporting period but that was not yet received. The time series of incoming quantity shortages thus comprises units that are still in transit at the time of reporting or for which there are longer delays in delivery. For most SKUs, this time series predominantly contains zeros and low values in proportion to total shipments to wholesalers. This indicates that a delay in delivery is the exception rather than the rule. Taking this into account, and since there are shipments to wholesalers in almost all periods for most SKUs, we only include one lag for the incoming quantity shortages. The missing values for the first 5 observations, resulting from the use of these lagged inputs, are dealt with by deleting these observations.

Depending on the forecast horizon h, not all future values of the lagged exogenous variables are known at the time when the forecast is to be generated. Two common strategies exist to deal with this problem: (i) predicting unknown variables separately resulting in the introduction of additional forecast errors; or (ii) restricting inputs to an unconditional or ex ante forecasting setup (Ord et al., 2017). In the unconditional forecasting setup, model inputs are restricted to only contain variables for which the future values are available, i.e., with k the lag order for an exogenous variable, only variables for which $k \geq h$ are included. For example, taking into account the lag order restrictions as described above, for forecast horizon $h = 5$, besides the seasonal dummies, only the variables 'wholesaler sales lag 5' and 'wholesaler ending inventory lag 5' are considered as model inputs. This unconditional forecasting setup thus implies that the forecasting model is reformulated for each forecast horizon h as a different constraint on k is imposed. Taking into account the forecastability of the time series included in the sell-through information, we opt for this second strategy.

We first make a selection of the exogenous variables before adding them to the forecasting models. As we rely on the unconditional forecasting setup, we use a two-stage selection approach. First, we select seasonal dummy variables regardless of the forecast horizon h following the procedure described in Section 4.2.1. Second, we also rely on the hybrid stepwise selection strategy for selecting sell-through information features. For this purpose, however, we replace manufacturer shipments (y) by the residuals of a non-seasonal ETS or ARIMA model (depending on the time series method considered) with the selected seasonal dummies modeled as exogenous variables. In this way, we emphasize that exogenous information is used as an additional instrument in the forecasting process, after having captured the dynamics in the time series itself. Note that although a stepwise selection strategy is a commonly used method to select among potentially useful variables, it has been criticized to result in high variance solutions; see e.g., Hastie et al. (2011).

To predict future demand with ETSX and ARIMAX we thus rely on prior shipments to wholesalers (y), the set of selected seasonal dummy variables D^*, and the set of selected sell-through variables X_h^*:

$$\hat{y}_{t+h|t} = f(y_t, y_{t-1}, \ldots, y_{t-m}, D^*, X_h^*). \tag{4}$$

Recall that X_h, the set of available sell-through variables, also varies with the forecast horizon h as we use an unconditional forecasting setup.

Machine learning techniques – The seasonal and sell-through information variables discussed above are complemented by five autoregressive (AR) lags and a linear trend variable to form the model inputs for the ML methods. This enables us to capture the time series dynamics itself. For the maximum lag orders of the sell-through variables, we use the same values as specified above for the extrapolative time series methods with exogenous variables.

Also for the ML methods, we rely on the unconditional forecasting setup as discussed above. As this approach implies that the forecasting models are reformulated for each forecast horizon h, we make a selection of the seasonal dummy variables following the procedure described in Section 4.2.1 to ensure that the set of selected seasonal variables remains constant across forecast horizons. The four ML methods considered all contain regularization mechanisms to prevent overfitting, hence, all available (lagged) variables (sell-through information, AR terms and trend) are used as model inputs. The AR terms are important in capturing the shorter-term dynamics because the minimum lag order is not affected by the forecast horizon h as is the case for the set of available sell-through variables X_h. Indeed, we can use the shorter-horizon forecasted values as input without the need for a separate forecasting model, i.e., in predicting $t+5$ we can use the predictions for $t+1$ to $t+4$ as AR inputs.

In this empirical study we consider standard implementations of four commonly used ML techniques discussed in Section 3.2. More advanced variants of these methods do exist; however, we consider this to be out of scope for this study. The standard implementations for the ML techniques considered all require the specification of at least one hyperparameter. To this end, we perform grid search and use a $3\times$ repeated 10-fold cross-validation (CV) (Kuhn, 2008). This is a valid approach in a time series context when the model inputs include lagged values of the response variable (Bergmeir, Hyndman, & Koo, 2018). We rely on the software manuals (and the references therein) of the implementations used in order to specify the search grids used for hyperparameter tuning. More specifically, for each hyperparameter considered, the search grid contains a range of possible values centered around the default value. More details for each of the ML techniques are provided below.

The first ML technique is the LASSO regression (Hastie et al., 2011). This method aims to reduce the forecast variance by shrinking the parameter estimates in the traditional linear model via L1 regularization and as such performs simultaneously coefficient shrinkage and variable selection. To ensure that the LASSO regularization does not depend on the units of the model inputs, they are first standardized to have zero-mean and unit-variance. The LASSO technique contains one hyperparameter, the regularization parameter λ. A search grid of 100 λ-values is provided by creating a sequence of decreasing values from λ_{max} to $\lambda_{min} = 0.0001 * \lambda_{max}$ on the log scale and where λ_{max} is the smallest value of λ for which the LASSO selects no variables (Friedman, Hastie, & Tibshirani, 2010). The forecasts of the LASSO model are produced as follows:

$$\hat{y}_{t+h|t} = f(y_{t+h-1}, \ldots, y_{t+h-L}, T_{t+h}, D^*, X_h), \tag{5}$$

$$= \beta_0 + \sum_{l=1}^{L} \beta_l y_{t+h-l} + \beta_{L+1} T_{t+h} + \sum_{s=1}^{S} \beta_{L+1+s} d_s + \sum_{r=1}^{R} \beta_{L+1+S+r} x_{h,r}, \tag{6}$$

with L the maximum lag order, T_{t+h} the linear trend variable, β_i the regression coefficient for variable i, d_s the sth seasonal variable in set D^* with S the total number of selected seasonal dummies, and $x_{h,r}$ the rth sell-through variable in set X_h with R the total number of sell-through variables for horizon h. Each $x_{h,r}$ is a combination of a sell-through variable and a lag. The autoregressive terms y_{t+h-l} are substituted by their forecasts $\hat{y}_{t+h-l}$ if $l < h$. Note that use of the LASSO loss function may cause variables to be omitted from the model.

The second ML technique employed is an artificial neural network. For this study, we use the multilayer perceptron (MLP) architecture (Zhang et al., 1998). For most forecasting purposes, a single hidden layer is deemed sufficient because of the universal approximation property and empirical results reported in the literature. However, we model an MLP with both direct linear connections from the input layer to the output layer (called skip-layer connections), as well as through the nonlinear hidden layer, as explicitly formulating linear skip-layer connections can help the training of the model significantly (Venables & Ripley, 2002). In the hidden layer, we use the logistic activation function, while we use the linear activation function for the output layer. To enable the MLPs to detect and exploit the most important variables, model inputs are rescaled to the unit interval and an L2 regularization term is added to the loss function. The regularization parameter can take the value 0.001, 0.01 or 0.1. To determine the number of hidden nodes, there are no reliable rules of thumb. However, because we use an architecture with skip-layer connections, the nonlinear component can be regarded as an incremental addition to the linear model and therefore we restrict the number of hidden nodes to lie between 1 and 10 (with a step size of 1). Finally, the number of learning iterations is set to 1000.

A third ML technique that is used is support vector regression (SVR) (Smola & Schölkopf, 2004). SVR is a kernel method that performs nonlinear regression based on the kernel trick, implicitly mapping the inputs into high-dimensional feature spaces, while controlling the flatness of the regression function via L2 regularization. The aim is to find the function that has at most ϵ deviation from the actually obtained targets for all training data. To avoid overfitting, a cost parameter C controls the degree to which these constraints may be violated. Before training the model, inputs are standardized to have zero-mean and unit-variance. We use the radial basis function as kernel, which results in two hyperparameters that need to be specified: the cost parameter C and the kernel parameter σ. To identify good hyperparameter values we use exponentially growing sequences for $C = 2^{-5}, 2^{-4}, \ldots, 2^{15}$ and $\sigma = 2^{-15}, 2^{-14}, \ldots, 2^{3}$. The value for ϵ is kept constant at 0.1 (Karatzoglou, Smola, Hornik, & Zeileis, 2004).

The fourth ML technique considered in this study is random forests (RF) (Breiman, 2001). This ensemble method produces forecasts by combining the predictions of a large number of regression trees and uses both bagging and feature bagging to improve generalization performance. The RF method, like MLP and SVR, is able to capture nonlinear relations in the data. For each RF model, we grow 500 trees where sampling of observations is done with replacement. We tune two hyperparameters to control the randomness of the RF so as to achieve a reasonable strength of the single trees without too much correlation between the trees (bias-variance trade-off): the number of variables randomly sampled as candidates to possibly split at in each node and the minimal node size. The former can take on values between 2 and $2\lfloor p/3 \rfloor$, with a step size of 2 and p the number of model inputs, while the latter can take the values $1, 3, \ldots, 9$ (Wright & Ziegler, 2017).

For MLP, SVR and RF, forecasts are produced with the same inputs as in Eq. (5).

4.3. Evaluation

We simulate real-time forecasting by calculating out-of-sample forecasts at forecast horizons $h = 1$ to 5 weeks, which are relevant horizons to support the manufacturer's decision making. Recall from Section 4.1 that SKUs have a variable number of observations and sales of some products were discontinued during the observation period. Therefore, the time periods used to evaluate the out-of-sample forecasting performance are not fixed across SKUs: for each time series the last 30% of observations are withheld as a test set.

We adopt a rolling origin evaluation for each forecast horizon so that forecasts are produced for all weeks in the out-of-sample period by gradually increasing the in-sample period that is used for model fitting. In this way, multiple forecast errors can be used to evaluate the performance per horizon. Note that this implementation of rolling origin evaluation generates an equal number of forecast errors for each forecast horizon. For more details on rolling origin evaluation, one may refer to Tashman (2000).

The rolling origin evaluation is employed for two reasons: (i) to improve the reliability of the results as to enable proper evaluation of the methods, and (ii) to allow for changes in model parameters over time to counter potential concept drift. However, note that we restrict the latter to parameters only, as hyperparameters (for the ML techniques) are selected initially based on the shortest in-sample period. In the same regard, for selecting seasonal dummy variables and for the extrapolative time series methods with downstream exogenous variables, the variable selection procedure is only performed once.

To evaluate the forecast accuracy of the competing methods across SKUs, we use the Average Relative Root Mean Squared Error (AvgRelRMSE) (Davydenko & Fildes, 2013). The AvgRelRMSE_h is a scale-independent error measure that, for a given forecast horizon h, is based on the (weighted) geometric mean of the Relative RMSE of a forecasting method A over a baseline B across all SKUs:

$$\text{AvgRelRMSE}_h = \left(\prod_{i=1}^{m} \left(\frac{\text{RMSE}^{A}_{h,i}}{\text{RMSE}^{B}_{h,i}} \right)^{n_i} \right)^{1/\sum_{i=1}^{m} n_i}, \tag{7}$$

with n_i the number of periods in the out-of-sample test set for SKU i, m the total number of SKUs, and

$$\text{RMSE}_{h,i} = \sqrt{\frac{1}{n_i} \sum_{t \in T_i} (y_{i,t} - f_{i,t})^2}, \tag{8}$$

with $y_{i,t}$ and $f_{i,t}$ the actual and h-step ahead forecasted values at period $t \in T_i$, the set containing the time periods for SKU i for which out-of-sample forecasts are produced. Taking into account that our goal is to empirically evaluate the value of incorporating sell-through data in demand forecasting, we use ETS as baseline method.

The AvgRelRMSE is easy to interpret as it directly shows how a forecasting method improves or reduces the RMSE compared to the baseline method: the average percentage improvement in RMSE of forecasts is found as $(1 - \text{AvgRelRMSE}) \times 100$. Obtaining a value lower than one thus means that the forecasting method outperforms the baseline on average, while a value greater than one indicates the opposite. The AvgRelRMSE also has the following desirable properties: it is robust to outliers and calculation issues, and does not introduce biases (e.g., negative and positive forecast errors are penalized equally) which are common in e.g., percentage based error metrics. Finally, note that while we cannot rely on these percentage based error metrics due to periods in which manufacturer shipments are equal to zero, it is possible to use

Table 3
Forecast accuracy results - Average Relative RMSE across SKUs.

Method		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	*Overall*
NIS	Naïve	1.458	1.347	1.325	1.324	1.226	1.358
	ETS	1.000	1.000	1.000	1.000	1.000	1.000
	ARIMA	1.017	1.016	1.001	1.015	1.014	1.015
IS	ETS-W	0.970	0.983	0.987	0.988	0.987	0.984
	ARIMA-W	0.960	0.980	0.981	0.981	0.980	0.978
	ETSX	0.909	1.006	0.992	0.994	0.995	0.982
	ARIMAX	0.911	1.002	0.988	0.990	0.996	0.981
	LASSO	0.902	**0.966**	**0.962**	**0.961**	**0.961**	**0.953**
	MLP	1.010	1.079	1.082	1.099	1.062	1.078
	SVR	**0.900**	0.974	0.966	0.972	0.984	0.963
	RF	0.946	0.985	0.985	0.980	0.983	0.976
Impr. over best NIS		+10.0%	+3.4%	+3.8%	+3.9%	+3.9%	+4.7%

alternative scale-independent error measures based on scaled errors, such as the MASE (mean absolute scaled error) proposed by Hyndman and Koehler (2006). For more information, one may refer to Davydenko and Fildes (2013).

5. Empirical results

In this section, we assess whether the methods that include sell-through data, which represent the information sharing case (hereafter IS methods), are more accurate than those that do not (hereafter NIS methods). The results are first analyzed by out-of-sample forecast accuracy. Next, statistical tests are carried out to assess whether reported differences in forecasting performance are statistically significant. Finally, we discuss the reported forecast accuracy results in more detail and assess the effects of certain modeling choices.

5.1. Forecast accuracy

Forecast accuracy results of the empirical evaluation are provided in Table 3. For each method, we provide the AvgRelRMSE across SKUs for one- to five-step ahead forecasts and across forecast horizons. For each horizon, the best performing method is highlighted in boldface.

Considering the accuracy results in Table 3, various findings can be identified:

1. We find that ETS outperforms the other univariate baseline methods for all forecast horizons. The differences reported for the Naïve method indicate that using more advanced methods is certainly justified. For ARIMA, the reported decrease in forecast accuracy shows that it performs on average 1.5% worse than the baseline ETS method across the forecast horizons considered.
2. Comparing NIS and IS methods, we observe that there are IS methods that outperform the former for all forecast horizons considered. This indicates that using sell-through data can reduce forecast errors on average. SVR shows the best performance for $h=1$, while the LASSO has lowest RMSE compared to the baseline for $h>1$ and only shows a small difference in accuracy for $h=1$ compared to SVR. We further observe that both substitution methods, ETS-W and ARIMA-W, and LASSO, SVR and RF, outperform the NIS methods for all horizons. However, the results also indicate that there are substantial differences in the performance of IS methods both between and within forecast horizons.
3. For forecast horizons $h>1$, we find the accuracy of the substitution methods, ETS-W and ARIMA-W, to be much closer to that of the LASSO compared to the results for $h=1$. As the substitution methods only take the wholesaler sales into account, this observation suggests that in this case study, the information contained in wholesaler ending inventory positions (and incoming quantity shortages) is highly valuable for $h=1$ but becomes less important for $h>1$. As we use the same time series methods for the substitution methods and the NIS baselines, with the difference being that the former rely on wholesaler sales (w) instead of manufacturer shipments (y) as input, the improvements in forecast accuracy observed for the substitution methods thus result from the smoother input signal being used (see Fig. 2).
4. The performance of the extrapolative time series methods with exogenous variables, ETSX and ARIMAX, is very close to that of SVR and LASSO for $h=1$. However, for longer forecast horizons, the improvements in accuracy compared to the baseline ETS method are limited.
5. While LASSO, SVR and RF outperform the baseline ETS method for all horizons, MLP consistently performs worse. Although SVR does slightly better than LASSO for $h=1$, we can conclude that LASSO is the best performing ML technique as it dominates SVR for $h>1$. Further, while the performance of RF for $h=1$ lies between that of the substitution methods and that of LASSO and SVR, for $h>1$ it is comparable to that of the substitution methods. Recall that MLP, SVR and RF allow modeling of nonlinear relations in the data. These findings seem to indicate that allowing for nonlinear relations leads to a deterioration in forecast accuracy, though in varying degrees.
6. Considering the percentage improvements of the best IS methods over the best NIS method shows that we can achieve an average improvement of 4.7% in RMSE compared to the baseline ETS method across the forecast horizons considered. Forecast horizon specific percentage improvements indicate that the value of sell-through data is highest for $h=1$ with an average improvement in RMSE of 10% over the baseline. For longer forecast horizons, the reported percentage improvements are substantially smaller and lie within the range of 3.4% to 3.9%. Note that this finding is in line with previous results reported in the literature (see Section 2.2).

5.2. Statistical tests

In this section, we report the results of statistical tests that we performed to assess whether the reported differences in accuracy are statistically significant. We follow the procedure described in Davydenko and Fildes (2016) and use the non-parametric Friedman test, which is analogous to the well-known repeated-measures ANOVA test without assumptions of normality (Friedman, 1940). The Friedman test is based on comparing the mean rank of each method, testing if there are some significant differences between the methods.

Table 4
Mean ranks and associated p-values for the Bonferroni-Dunn post-hoc test for Relative RMSEs.

Method		$h = 1$		$h = 2$		$h = 3$		$h = 4$		$h = 5$		*Overall*	
NIS	Naïve	10.56	$>.999$	10.52	$>.999$	10.78	$>.999$	10.64	$>.999$	10.66	$>.999$	10.92	$>.999$
	ETS	7.39	–	5.78	–	5.58	–	5.73	–	5.59	–	6.27	–
	ARIMA	7.07	$>.999$	5.68	$>.999$	6.10	$>.999$	6.09	$>.999$	5.83	$>.999$	6.39	$>.999$
IS	ETS-W	6.24	0.415	5.36	$>.999$	5.64	$>.999$	5.56	$>.999$	5.62	$>.999$	5.68	$>.999$
	ARIMA-W	5.82	0.090	4.96	$>.999$	5.06	$>.999$	5.12	$>.999$	5.10	$>.999$	5.02	0.298
	ETSX	4.33	$<.001$	6.00	$>.999$	5.44	$>.999$	5.55	$>.999$	5.47	$>.999$	4.93	0.217
	ARIMAX	4.19	$<.001$	5.12	$>.999$	5.48	$>.999$	5.09	$>.999$	5.31	$>.999$	4.69	0.086
	LASSO	**3.40**	$<.001$	**3.94**	0.028	**3.78**	0.033	**3.58**	0.006	**3.64**	0.016	**3.16**	$<.001$
	MLP	7.20	$>.999$	8.48	$>.999$	8.30	$>.999$	8.66	$>.999$	8.40	$>.999$	8.92	$>.999$
	SVR	4.06	$<.001$	4.68	0.486	4.34	0.308	4.72	0.639	4.60	0.678	4.40	0.024
	RF	5.74	0.064	5.48	$>.999$	5.50	$>.999$	5.26	$>.999$	5.78	$>.999$	5.62	$>.999$

The Friedman tests for the Relative RMSE figures conclude that there are significant differences between the methods with all reported p-values less than 0.001. Here we proceed with a post-hoc test to find out which methods actually differ. As our research objective is to assess whether sell-through data can help in improving the forecast accuracy, the post-hoc test is used to identify whether the different methods considered perform significantly better than the best NIS method, i.e., the ETS baseline. To compare all methods to a control, we can rely on the Bonferroni-Dunn test. For more information, one may refer to Demšar (2006).

Mean ranks, along with their associated p-values, are provided in Table 4. A mean rank of one represents a method being the best for every single case, while a rank of 11 indicates that the method is always the worst. An associated p-value less than 0.05 indicates that the mean rank is significantly lower than that of the ETS baseline under 5% significance level, which is the significance level that we adopt. LASSO achieves the best ranking for all forecast horizons separately and across forecast horizons. Moreover, its mean rank is always significantly lower than that of the ETS baseline. Focusing on the results per forecast horizon, the mean ranks and the reported p-values indicate that the sell-through data is highly valuable for horizon $h = 1$, where all explanatory methods rank better than the NIS baselines, except for the poorly performing MLP. However, also here, we observe from the reported mean ranks that the improvements in accuracy for the IS methods are substantially smaller for $h > 1$ with only the LASSO mean ranks being significantly lower than the mean ranks of the ETS baseline.

5.3. Discussion

To assess the findings from the previous sections in more detail, for all forecast horizons, we visualize both forecast accuracy and mean rank results in Fig. 4 for the following selection of methods: ETS, ARIMA-W, ARIMAX, LASSO and SVR. The left panel displays the AvgRelRMSE for each forecast horizon. In the right panel, for the same methods, we visualize the evolution of the mean ranks for the Relative RMSE figures with respect to the forecast horizon.

Based on the forecast accuracy and mean rank results, we observe that the selected IS methods generally outperform the NIS ETS baseline. We can thus conclude that the use of sell-through data allows the manufacturer to effectively improve its short-term forecast accuracy. We can further conclude that LASSO is the best performing method in our case study as it generally outperforms the other methods both in terms of AvgRelRMSE and mean rank. The fact that LASSO also outperforms SVR indicates that allowing for nonlinear relations in the data is not essential in this case.

In Fig. 4, both panels clearly illustrate a difference between $h = 1$ and $h > 1$, with differences in accuracy between the IS methods and the ETS baseline being substantially greater for $h = 1$. For ARIMAX, the performance for $h = 1$ is close to that of SVR and LASSO; however, for $h > 1$ there is only limited improvement in accuracy compared to the baseline ETS method. In terms of both the AvgRelRMSE and mean rank, for horizon $h = 1$, we further observe a difference between LASSO and SVR on the one hand and ARIMA-W on the other hand. However, for longer horizons $h > 1$, the differences in AvgRelRMSE and mean ranks between these two groups become smaller. This observation can likely be explained by the following two elements. First, for longer forecast horizons, less sell-through information is made available to the explanatory methods as we only allow for higher lag orders of this information due to the unconditional forecasting setup used. Specifically, the information on incoming quantity shortages is only available for horizon $h = 1$. Second, as discussed in Section 5.1 in light of the performance of both substitution methods for $h > 1$, the higher order lags of the wholesaler ending inventory positions do not seem to exhibit strong predictive information as compared to the lowest order lag.

Both elements and their implications become apparent when analyzing the inputs that are selected by LASSO for the shortest in-sample period. A visualization is provided in Fig. 5, where selected variables are labeled in blue. Darker blue indicates a greater weight attached to the selected variable[6] and unavailable variables are colored in gray. For horizon $h = 1$, LASSO clearly selects more variables, with an average of 12 variables, while for horizons $h > 1$ only 7 variables are selected on average. We further observe that the variables ‘Sales lag 1’ and ‘Inventory lag 1’ often seem to have strong predictive value for horizon $h = 1$, while rather small weights are assigned to the variable ‘INC QTY shortage lag 1’ as expected. As the aforementioned variables cannot be selected for horizons $h > 1$, LASSO often puts more weight on ‘Sales lag 5’ and the forecasted AR lags enter the models. As these forecasted AR lags are based on shorter-horizon forecasted values, LASSO tries to capture the sell-through information implicitly. For forecast horizons $h > 1$, information regarding the ending inventory positions of the wholesaler seems to have limited predictive power on average as these variables are selected for only a few SKUs. Finally, note that the selection of variables can change across forecast horizons, even after we take into consideration the availability of the variables. This is because LASSO considers interactions between possible input variables in making the selection, which can prevent available inputs selected for a specific forecast horizon from being selected for other horizons.

We set up two extra experiments to investigate more thoroughly the performance of LASSO. In the first experiment, in order to separate the effects of forecasting method and sell-through data on forecasting performance, we compare the LASSO model, as detailed in Section 4.2.2, with a LASSO-NIS model in which no sell-through information features are available to predict future demand. We also add a linear regression (LR) model for both the NIS

[6] Coefficients are multiplied by the standard deviation of their respective variable to visualize the importance of the selected variables.

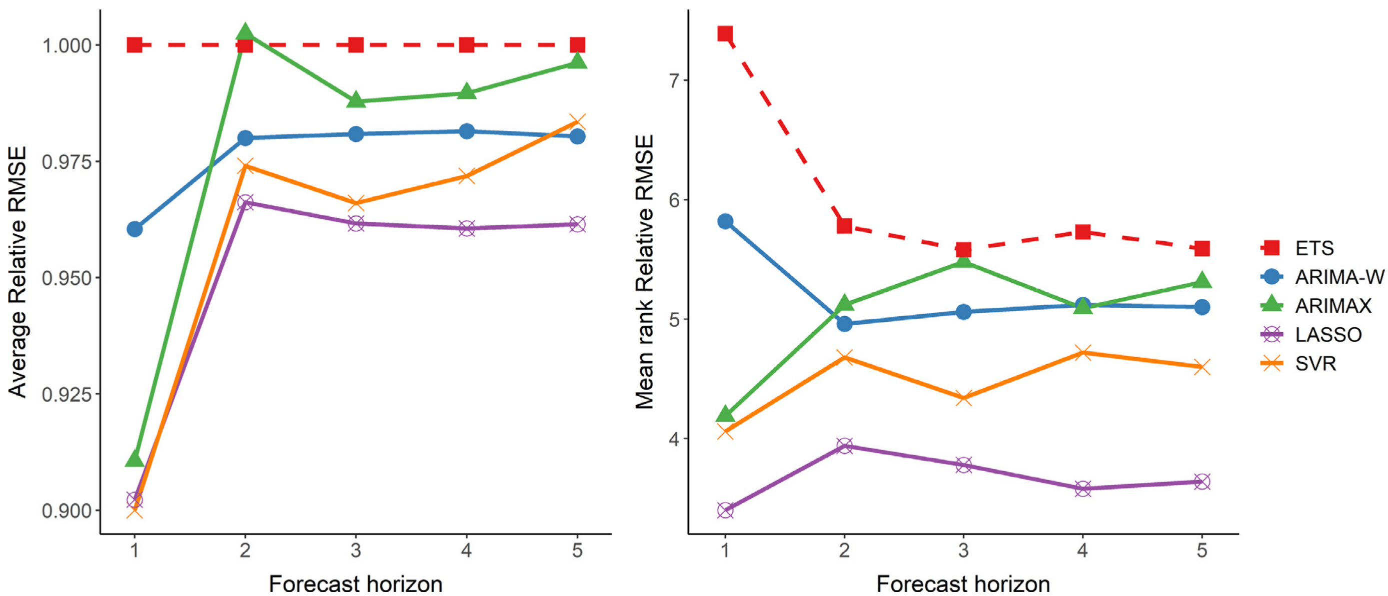

Fig. 4. Average Relative RMSE and mean ranks for Relative RMSEs.

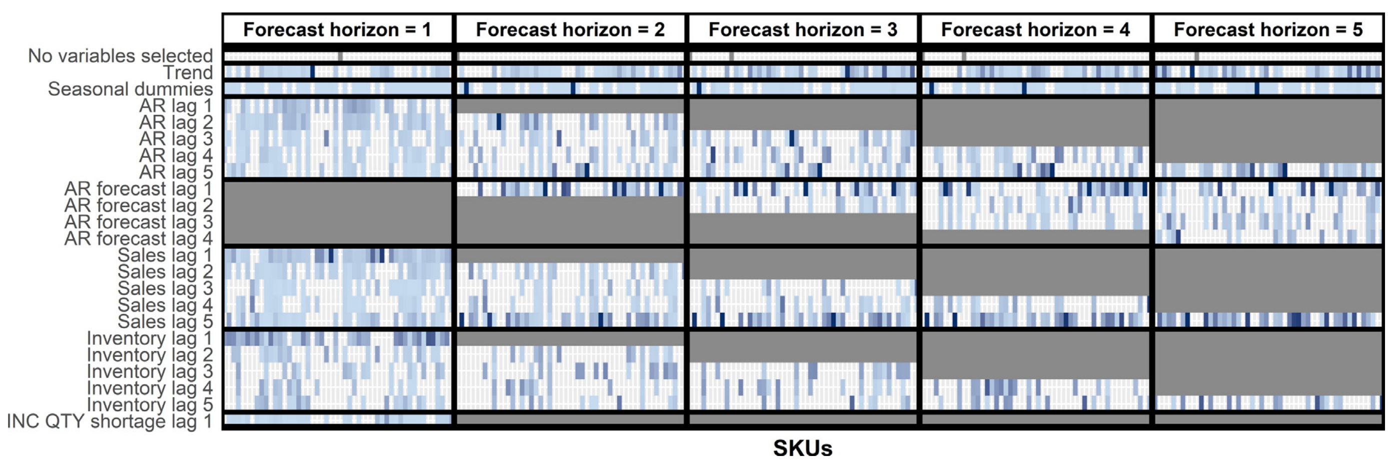

Fig. 5. Heatmap of variables selected by LASSO per forecast horizon.

Table 5
Separating the effects of forecasting method and sell-through data in terms of Average Relative RMSE.

Method	Type	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	*Overall*
ETS	NIS	1.000	1.000	1.000	1.000	1.000	1.000
ARIMA-W	IS	0.960	0.980	0.981	0.981	0.980	0.978
LR	NIS	0.968	0.995	0.992	0.988	0.980	0.985
	IS	0.884	0.995	0.983	0.981	0.982	0.969
	Diff.	*+0.084*	*+0.000*	*+0.009*	*+0.007*	*-0.002*	*+0.016*
LASSO	NIS	0.953	0.982	0.975	0.975	0.971	0.972
	IS	0.902	0.966	0.962	0.961	0.961	0.953
	Diff.	*+0.051*	*+0.016*	*+0.013*	*+0.014*	*+0.010*	*+0.019*

and IS setting. For these LR models, the hybrid stepwise selection strategy as outlined in Section 4.2.1 is used for variable selection. A two-stage approach is adopted here as to keep the selected seasonal dummy variables fixed across forecast horizons. In a second experiment, we assess the sensitivity of LASSO with regard to the number of model inputs.

The results of the first experiment are shown in Table 5. These results effectively allow us to separate the effects of forecasting method and sell-through data on forecasting performance. The AvgRelRMSE figures for LASSO-NIS clearly indicate that a LASSO model without sell-through information features also outperforms the ETS baseline for all forecast horizons. This allows us to conclude that only part of the reported decrease in RMSE for LASSO compared to the ETS baseline originates from the use of sell-through information. For $h=1$ in particular, only half of the reduction in AvgRelRMSE for LASSO can be attributed to the use of sell-through information. Recall from Section 5.1 that the improvements in forecast accuracy observed for the substitution methods result from the smoother wholesaler sales input signal. Comparison of LASSO-NIS and ARIMA-W results seems to indicate that LASSO-NIS can achieve similar performance without the use of the smoother input signal. For the LR model without sell-through information, we also observe an improvement in accuracy compared to the ETS baseline, but it is outperformed by ARIMA-W. While LASSO benefits from the addition of sell-through information for all forecast horizons, this is not the case for the LR model, where it mainly affects the accuracy for $h=1$. Interestingly, for $h=1$ the LR model with sell-through data even outperforms LASSO.

Table 6 depicts the results of the second experiment in which we assess the sensitivity of LASSO with regard to the number of model inputs by increasing the maximum lag order from 5 to 10 for the AR terms and the wholesaler sales and ending inventory position features. The analysis is also performed for ARIMAX. Re-

Table 6
Effect of maximum lag order (MLO) on Average Relative RMSE.

Method	MLO	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	*Overall*
ETS	N/A	1.000	1.000	1.000	1.000	1.000	1.000
ARIMAX	5	0.911	1.002	0.988	0.990	0.996	0.981
	10	0.920	1.017	1.001	0.999	0.996	0.990
	Diff.	*−0.009*	*−0.015*	*−0.013*	*−0.009*	*+0.000*	*−0.009*
LASSO	5	0.902	0.966	0.962	0.961	0.961	0.953
	10	0.905	0.985	0.982	0.983	0.984	0.971
	Diff.	*−0.003*	*−0.019*	*−0.020*	*−0.022*	*−0.023*	*−0.018*

Table 7
Effect of number of hidden nodes (NHN) in MLP on Average Relative RMSE.

Method	NHN	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	*Overall*
ETS	N/A	1.000	1.000	1.000	1.000	1.000	1.000
LASSO	N/A	0.902	0.966	0.962	0.961	0.961	0.953
MLP	1–10	1.010	1.079	1.082	1.099	1.062	1.078
	0–10	0.915	0.999	0.994	0.992	0.987	0.980
	Diff.	*+0.095*	*+0.080*	*+0.088*	*+0.107*	*+0.075*	*+0.098*

call that for this method, we rely on a hybrid stepwise selection strategy for variable selection, while the LASSO relies on shrinkage. For the LASSO we observe that forecasting performance deteriorates as more model inputs are made available. However, the impact is rather limited for $h=1$. The results for ARIMAX show a similar pattern, further deteriorating the accuracy for $h>1$. The results thus indicate that in our case both shrinkage and the hybrid stepwise selection strategy suffer from the addition of extra model inputs for $h>1$. As these observations suggest that the leading effects in the sell-through data are more easily identified (and as such are more pronounced) for $h=1$, they reinforce our conclusion that the value of sell-through data is highest for this forecast horizon.

Recall from Fig. 4 that LASSO also outperforms SVR, indicating that allowing for nonlinear relations in the data is not essential in this case. This observation may also explain the poor performance of MLP as the architecture as outlined in Section 4.2.2 always adds at least one nonlinear hidden node to the linear component of the model (we use an architecture with linear skip-layer connections), with the number of hidden nodes determined via CV. In Table 7, we show the results from a comparison of this MLP architecture with an alternative one where the number of hidden nodes can lie between 0 and 10. We thus allow the MLP to collapse into a fully linear model. The results show improvements in accuracy over the nonlinear MLP architecture between 7.5 and 10.7 percentage points, clearly indicating that the enforcement of a nonlinear component harms the forecast accuracy. However, note that also the linear MLP is dominated by LASSO for all forecast horizons.

Based on the reported findings, we argue that sell-through data can effectively be used to improve the forecast accuracy at manufacturer level. We find evidence for an improvement in forecast accuracy for all horizons considered, with the greatest improvement observed for one-step ahead forecasts. A profound analysis of the LASSO results provides insight into why this is the case: the lowest order lags seem to contain the most predictive information. The latter may be explained by the short delivery lead times in our case study, with the delivery lead time being the time between the placement of an order by a wholesaler and the moment of delivery of this order. Recall that the BWE implies that there is an information lead time, i.e., orders received by wholesalers reflect changes in final customer demand earlier than orders received by the manufacturer. Short delivery lead times, however, do not incentivize wholesalers to anticipate potential changes in final customer demand multiple periods ahead which causes them to react quickly to the perceived demand. In this regard, in other multi-echelon supply chains with longer delivery lead times we would possibly observe greater forecast accuracy improvements for longer forecast horizons.

6. Conclusions

In this paper, we consider the use of shared sell-through data as a source of downstream information in multi-echelon supply chains to improve the accuracy of short-term demand forecasts at manufacturer level. The academic literature on information sharing as a means to counter the impact of the BWE is equivocal. However, the limited number of data-driven empirical studies mostly provide evidence on the benefits of using downstream information in demand forecasting, especially for short forecast horizons. This work extends the existing empirical studies by considering sell-through data originating from intermediaries instead of POS data.

In line with previous results on the use of POS data, this paper provides empirical evidence of the value of sell-through data to improve forecast accuracy at manufacturer level and as such provides indirect evidence that its use allows mitigating the impact of the BWE. The reported empirical findings result from a case study involving a real dataset including 50 different SKUs from a US drug manufacturer operating in a multi-echelon supply chain. For this case study, we consider forecast horizons $h=1$ to 5 and apply both time series methods and machine learning techniques. Moreover, where previous studies solely focused on downstream demand as such, we also consider information on wholesaler inventory positions (and incoming quantity shortages). The results point to LASSO as best method and provide evidence of an improvement in forecast accuracy for all horizons considered, with the greatest improvement observed for one-step ahead forecasts. Analysis of the variables selected by LASSO clarifies that the latter results from the fact that the lowest order lags seem to contain the most predictive information, which may be explained by the short delivery lead times for the manufacturer in our case study. In this regard, we can conclude that potential accuracy gains in other multi-echelon supply chains may depend on the characteristics of the involved supply chain, and more specifically on the prevailing delivery lead times.

Although additional empirical studies on the use of sell-through data are needed to more broadly confirm our results, there seems to be a clear potential and hence an incentive for manufacturers to assess whether they can achieve accuracy improvements by using sell-through data. This is true for all manufacturers operating in multi-echelon supply chains, and in particular if POS data is not (timely) available for short-term forecasting purposes as in the presented case study. At the same time, this lack of (timely) available POS data is also a limitation of our empirical study, as in theory both downstream information sources, POS and sell-through data, could be used by manufacturers operating in multi-echelon supply chains. In this regard, an interesting topic for future research is an overarching empirical study which takes both sources of downstream information into account. Such an experimental setup would allow us to investigate whether the simultaneous use of both downstream information sources can further enrich the forecasting process and evaluate the incremental added value of both data sources. Moreover, it may enable us to identify several factors that affect POS and sell-through data differently, resulting in differences in their degree of usefulness. Two obvious examples of such factors are frequent stockouts and promotional activities at retailer level, which can both cause the sell-through data to be more useful to the manufacturer compared to POS data (also see Section 1). However, an experimental setup including both data sources may possibly lead to the identification of more subtle factors.

Acknowledgments

This research is funded by Agentschap Innoveren & Ondernemen Vlaanderen (project HBC.2016.0589) and OM Partners NV. We also would like to thank three anonymous reviewers for their insightful comments that helped us substantially improve the paper.

References

Ali, Ö. G., Sayın, S., Van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications, 36*(10), 12340–12348.
Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management, 29*(1), 19–31.
Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis, 120*, 70–83.
Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks.
Byrne, P., & Heavey, C. (2006). The impact of information sharing and forecasting in capacitated industrial supply chains: A case study. *International Journal of Production Economics, 103*(1), 420–437.
Byrne, R. F. (2012). Beyond traditional time-series: Using demand sensing to improve forecasts in volatile times.. *Journal of Business Forecasting, 31*(2).
Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science, 46*(3), 436–443.
Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting, 29*(3), 510–522.
Davydenko, A., & Fildes, R. (2016). Forecast error measures: Critical review and practical recommendations. In M. Gilliland, L. Tashman, & U. Sglavo (Eds.), *Business Forecasting: Practical Problems and Solutions*. John Wiley & Sons.
Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*(Jan), 1–30.
Di Pillo, G., Latorre, V., Lucidi, S., & Procacci, E. (2016). An application of support vector machines to sales forecasting under promotions. *4OR, 14*(3), 309–325.
Forrester, J. (1961). *Industrial dynamics*. Cambridge: M.I.T. Press.
Fransoo, J. C., & Wouters, M. J. (2000). Measuring the bullwhip effect in the supply chain. *Supply Chain Management: An International Journal, 5*(2), 78–89.
Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22.
Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics, 11*(1), 86–92.
Gartner Inc. (2019). It glossary. Accessed 5 April 2019https://www.gartner.com/it-glossary/.
Hanssens, D. M. (1998). Order forecasts, retail sales, and the marketing mix for consumer durables. *Journal of Forecasting, 17*(3–4), 327–346.
Hartzel, K. S., & Wood, C. A. (2017). Factors that affect the improvement of demand forecast accuracy through point-of-sale reporting. *European Journal of Operational Research, 260*(1), 171–182.
Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The elements of statistical learning: Data mining, inference, and prediction* (2nd). NY: Springer.
Holweg, M., Disney, S., Holmström, J., & Småros, J. (2005). Supply chain collaboration: Making sense of the strategy continuum. *European Management Journal, 23*(2), 170–181.
Hosoda, T., Naim, M. M., Disney, S. M., & Potter, A. (2008). Is there a benefit to sharing market sales information? Linking theory and practice. *Computers & Industrial Engineering, 54*(2), 315–326.
Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research, 237*(2), 738–748.
Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, Articles, 27*(3), 1–22.
Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.
Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin: Springer-Verlag.
Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab – an S4 package for kernel methods in R. *Journal of Statistical Software, 11*(9), 1–20.
Kelepouris, T., Miliotis, P., & Pramatari, K. (2008). The impact of replenishment parameters and information sharing on the bullwhip effect: A computational study. *Computers & Operations Research, 35*(11), 3657–3670.
Kim, K. K., Ryoo, S. Y., & Jung, M. D. (2011). Inter-organizational information systems visibility in buyer-supplier relationships: The case of telecommunication equipment component manufacturing industry. *Omega, 39*(6), 667–676.
Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics, 181*, 145–153.
Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26.
Lee, H., Kim, M. S., & Kim, K. K. (2014). Interorganizational information systems visibility and supply chain performance. *International Journal of Information Management, 34*(2), 285–295.
Lee, H. L., Padmanabhan, V., & Whang, S. (1997a). The bullwhip effect in supply chains. *SLOAN Management Review, 38*(3), 93–102.
Lee, H. L., Padmanabhan, V., & Whang, S. (1997b). Information distortion in a supply chain: The bullwhip effect. *Management Science, 43*(4), 546–558.
Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research, 249*(1), 245–257.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE, 13*(3), e0194889.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting, 36*(1), 54–74.
Ord, K., Fildes, R., & Kourentzes, N. (2017). *Principles of business forecasting* (2nd). Wessex Press Publishing Co.
Raghunathan, S. (2001). Information sharing in a supply chain: A note on its value when demand is nonstationary. *Management Science, 47*(4), 605–610.
Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., & Desmet, B. (2018). Tactical sales forecasting using a very large set of macroeconomic indicators. *European Journal of Operational Research, 264*(2), 558–569.
Sanders, N. R., & Graman, G. A. (2009). Quantifying costs of forecast errors: A case study of the warehouse environment. *Omega, 37*(1), 116–125.
Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222.
Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics – Theory and Methods, 7*(1), 13–26.
Svetunkov, I. (2020). Smooth: Forecasting Using State Space Models. R package version 2.5.6. https://CRAN.R-project.org/package=smooth.
Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1–26.
Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting, 16*(4), 437–450.
Trapero, J. R., Kourentzes, N., & Fildes, R. (2012). Impact of information exchange on supplier forecasting performance. *Omega, 40*(6), 738–747.
Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society, 66*(2), 299–307.
Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting, 29*(2), 234–243.
Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th). New York: Springer.
Williams, B. D., & Waller, M. A. (2010). Creating order forecasts: Point-of-sale or order history? *Journal of Business Logistics, 31*(2), 231–251.
Williams, B. D., & Waller, M. A. (2011). Top-down versus bottom-up demand forecasts: The value of shared point-of-sale data in the retail supply chain. *Journal of Business Logistics, 32*(1), 17–26.
Williams, B. D., Waller, M. A., Ahire, S., & Ferrier, G. D. (2014). Predicting retailer orders with POS and order data: The inventory balance effect. *European Journal of Operational Research, 232*(3), 593–600.
Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software, 77*(1), 1–17.
Zhang, C., Tan, G.-W., Robb, D. J., & Zheng, X. (2006). Sharing shipment quantity information in the supply chain. *Omega, 34*(5), 427–438.
Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62.