ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230511175500

Preliminary communication
Received: 2023-05-11, Accepted: 2023-05-15

A Dynamic Systems Model for an Economic Evaluation of Sales Forecasting Methods

Lara Kuhlmann*, Markus Pauly

Abstract: Sales forecasts are essential for a smooth workflow and cost optimization. Usually, they are assessed using statistical error measures, which might be misleading in a business context. This paper proposes a new dynamic systems model for an economic evaluation of sales forecasts. The model describes the development of the inventory level over time and derives the resulting overstock and shortage costs. It is tested on roughly 3,000 real-world time series and compared with the commonly used approach based on statistical measures. The experiments show that different statistical measures have no coherent evaluation, making their usage even less suitable for a practical economic application.

Keywords: forecasting; inventory management; sales forecast; supply chain analytics; time series

1 INTRODUCTION

Working with precise sales forecasts is crucial for the supply chains of production companies or retailers [1]. It enables a smooth workflow and reduces waste along the value chain [2]. The longer the lead times of products are, the more important it is to accurately plan ahead [3]. Due to current crises, the lead times of most products have increased substantially [4], amplifying the relevance of reliable forecasts. The challenge is to find the most suitable forecasting method among plenty of existing ones, each with its own benefits [5]. Usually, a forecasting method is chosen based on statistical accuracy measures, also called error metrics.

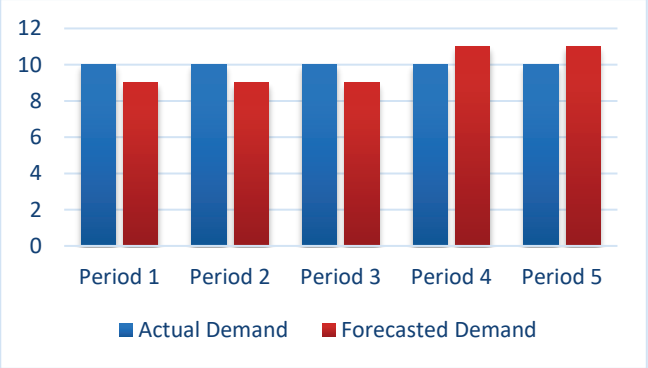

Figure 1 Forecasting with Model 1 in the toy example

However, solely choosing a sales forecast method based on statistical accuracy measures can have some disadvantages, as the following toy example illustrates: We assume a given demand for a product and we want to compare two forecasting models that forecast the demand for five periods. The forecast of Model 1 is one unit below the actual demand for the first three periods and one unit above the actual demand for the last two periods (see Fig. 1). The forecast of Model 2 is alternating one unit above and then one unit below the actual demand (see Fig. 2). Both models had the same absolute deviations in every period. Thus, statistical measures as the Root Mean Square Error (RMSE) or the Mean Average Percentage Error (MAPE) are the same for both models, suggesting an equally good performance.

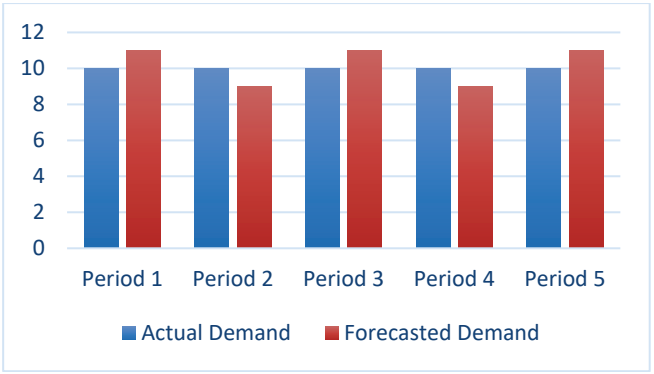

Figure 2 Forecasting with Model 2 in the toy example

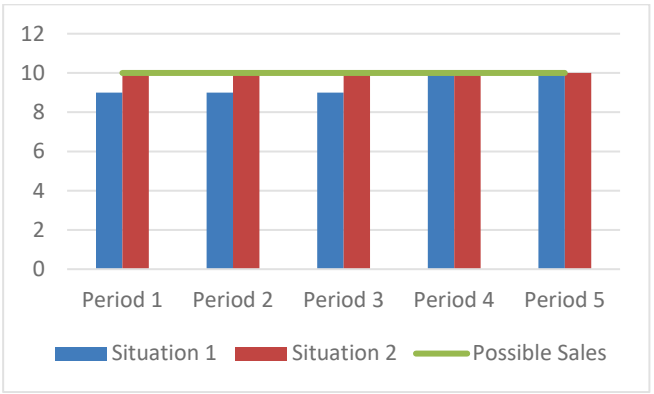

Figure 3 Sales based on the chosen forecasting Models 1 and 2 in the toy example

Nevertheless, from a business perspective, Model 2 outperforms Model 1 if the goods are non-perishable and their value does not decrease over time. If the company had acted upon Model 1, it would not have been able to meet the customers' demand in the first three periods (see Fig. 3). If the company had relied upon Model 2, it would have satisfied the customers' demand in every period. Thus, considering the costs of product shortage (which would have been higher for Model 1) and the costs of overstock (which would have been

equal) over time seems to be a more suitable way for evaluating sales forecasts.

Some papers already considered costs in the context of forecasting, but never the development of stock levels over time. In order to fill this gap, we introduce an intuitive dynamic systems model for an economic evaluation of sales forecasting methods, considering overstock and shortage costs that derive from the development of inventory levels. The new method is tested on a real-world dataset of about 3,000 product sales time series of a German raw materials wholesaler. Sales forecasts are created and evaluated using the dynamic systems model but also statistical measures. The evaluations are compared with regard to the similarity and preference for a certain forecasting method.

2 RELATED WORK

2.1 Sales Forecasting Methods

There are many different methods for sales forecasting [5]. They rely on time series analysis and forecast future demand based on historic sales [6]. External data can be integrated into forecasts to improve them [7]. The forecasting models are usually univariate point forecasts. The models can either be statistical time series models or machine learning models. The statistical models are easy to implement and intuitively explainable. Among what [8] considers classical sales forecasting methods are ARIMA (Autoregressive Moving Average) and Holt-Winters. As the name suggests, the ARIMA model combines an autoregressive component, that links past and present values in a similar way as autocorrelation is computed, and a moving average component [9]. Holt-Winters is a seasonal smoothing method that can capture both trends and seasonal behavior within a time series [10, 11]. Recently, a lot of research has been done on forecasting with machine learning methods [12]. They can provide forecasts with higher accuracy [13], but require more run time and are not intuitively explainable. However, as the no-free lunch theorem suggests, there is no single best method [14]. Depending on the characteristics of a dataset, one method achieves more precise forecasts than another method. Subsequently, it is advisable to test several forecasting algorithms and choose the most suitable one.

2.2 Statistical Sales Forecast Evaluation

The performance of sales forecasts can be evaluated by applying point forecast error metrics to a test dataset or multiple test datasets in case of (rolling) time series cross-validation. Ref. [15] conducted a survey on forecast error measures and found that 23 different measures are in use (see Tab. 1). A list providing the full names of the error measures, whose acronyms are given in the table, can be found in the appendix.

In the context of sales forecast, the most commonly applied error measures are the RMSE [13, 8, 16, 17, 18], the MAPE [17, 18, 19] and MAE (Mean Absolute Error) [16, 18, 20]. The MAPE is easy to interpret but cannot be computed if the time series contains zeros [21]. As a remedy, its symmetric version sMAPE (Symmetric Mean Absolute Percentage Error) can be applied. It has the further advantage of penalizing under-forecasts more severely than over-forecasts. As the time series for our experimental evaluation (see Section 4) contain zeros, we focus on the three error measures RMSE, MAE and sMAPE. The error measures are computed based on the actual sales y_t and the forecasted sales $\hat{y}_t$ at time t [15]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_t - \hat{y}_t)| \quad (2)$$

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n} 200 * \left|\frac{(y_t - \hat{y}_t)}{(y_t + \hat{y}_t)}\right|. \quad (3)$$

Table 1 Forecasting error measures based on [15]

Category	Error measures
Absolute forecasting errors	MAE, MdAE, MSE, RMSE
Measures based on percentage errors	MAPE, MdAPE, RMSPE, RMdSPE
Symmetric errors	sMAPE, sMdAPE, msMAPE
Measures based on relative errors	MRAE, MdRAE,
Scaled errors	MASE, RMSSE
Relative measures	RMAE, RRMSE, LMR, PB(MAE)
Other error measures	mRMSE, inRSE, Std_AE, Std_APE

Despite the challenge of choosing an error measure, the problem arises that none of them can provide us with a unique best forecasting method. What is an excellent RMSE value for one forecasting method in one dataset, might be a mediocre value for another dataset. Depending on the demand stability, some products are easier to forecast than others. Thus, to achieve an objective evaluation of a forecasting method, it is recommendable to create a baseline forecast, to which other more advanced methods can be compared to in the process of benchmarking [5]. A very common baseline method is naïve forecasting, which simply assumes the future sales to be equal to the most recently observed sales [22].

2.3 Inventory-Related Costs

Holding inventory entails different kinds of costs. In order to minimize the overall costs, the different kinds need to be balanced. Literature on reducing inventory-related costs focuses on capital commitment costs, order costs, and shortage costs. Sometimes, costs for the inventory control system are also considered [23]. The capital commitment costs are usually priced with the weighted average cost of capital (WACC) [24]. Storage costs can usually be disregarded because most companies own their warehouses, making the storage costs fixed [23]. Only if companies are charged fees based on a variable level of inventory the companies should consider these fees when deciding on inventory order politics [24]. Shortage costs measure the costs that occur from not being able to sell a product. These include the lost profit margin and also reputational damage and a decrease in customers' loyalty [25].

2.4 Cost-Considering Sales Forecast Evaluation

The evaluation of sales forecasts with costs has only been done rarely. The authors of [26] focused on reducing overstock costs with the help of statistical forecasting methods but disregarded shortage costs. Ref. [27] compared the sales forecasts of five neural networks in terms of supply chain-related costs. They derived the costs by looking at the deviations of the forecast and the actual demand for every period individually, not considering the development of inventory levels.

However, the question of how many goods to order to maximize the profit has been investigated a lot. The so-called newsvendor problem is a popular problem in logistics and many approaches to solving it have been published [28]. It also considers shortage costs and tries to optimize the order quantity [25] but it only regards one period. This is why it is not applicable in this paper.

3 THE DYNAMIC SYSTEMS MODEL

A dynamic systems model is an analytical model that describes the changes in a system over time [29]. The proposed dynamic systems model aims to describe the changes in inventory level and the resulting costs over time. To use the model as an economic evaluation for sales forecasting methods, some assumptions had to be made:

- Goods are non-perishable and their value does not decrease over time
- Storage costs are fixed costs
- Goods are ordered at the beginning of each period, order costs do not change
- The ordered goods are delivered at the beginning of period $t + LT$ (lead time)
- The lead time per product is fixed
- The costs for holding a safety stock are calculated into the product price, thus only costs for an inventory level above the safety stock are considered overstock costs.

Ref. [30] reviewed and compared deterministic and statistical methods for the calculation of a safety stock. They found the statistical methods to be more accurate. Thus, we rely on the statistical method, for which the aforementioned assumptions hold, e.g. that the lead time is fixed. The dynamic systems model determines the safety stock s based on a service factor Z, which depends on the desired service level, the lead time l, and the standard deviation of the product's demand σ_y:

$$s = Z * \sqrt{l} * \sigma_y. \tag{4}$$

In the first step, the dynamic systems model describes the inventory level development, later the resulting costs. For every period t, the model calculates the inventory level at the end of that period (I_{E_t}) by subtracting the actual sales (y_t) from the inventory level at the beginning of that period (I_{B_t}). Naturally, an inventory level can never be negative. Thus, the formula to calculate I_{E_t} is given by

$$I_{E_t} = \max(I_{B_t} - y_t, 0). \tag{5}$$

Here, I_{B_t} depends on the inventory level at the end of the prior period ($I_{E_{t-1}}$) and the goods that had been ordered and are delivered at the beginning of each period (D_t) via

$$I_{B_t} = I_{E_{t-1}} + D_t. \tag{6}$$

The number of goods delivered equals the number of goods that have been ordered in a previous period. The time gap between order and delivery is determined by the lead time l of a product:

$$D_t = O_{t-l}. \tag{7}$$

The number of goods ordered depends on the predicted sales for the period in which the goods will be delivered ($\hat{y}_{t+l}$), the predicted sales for the period in which the goods are ordered ($\hat{y}_t$), the safety stock s and the inventory level at the moment of order (I_{B_t}):

$$O_t = \max(\hat{y}_{t+l} + s + \hat{y}_t - I_{B_t}, 0). \tag{8}$$

The derivation of this formula can be found in the Appendix. As the dynamic systems model aims to describe the inventory level development over a certain period of time, periods prior to $t = 0$ need to be modeled to calculate the number of goods that are delivered in $t = 0$. For the periods prior to $t = 0$ the following assumption holds:

$$D_{0-l} = \hat{y}_{t-l}. \tag{9}$$

Moreover, due to the mutual dependency of I_{B_t} and I_{E_t}, an assumption needs to be made about the initial inventory level:

$$I_{1_{0-l}} = s + \hat{y}_{t-l}. \tag{10}$$

Tab. 2 displays another toy example of how the development of the inventory level is calculated. The initial inventory level $I_{B_{-2}}$ is the sum of the predicted sales for the period $\hat{y}_{-2}$, which are 51, and the safety stock s, which is 634. The number of goods ordered in $t = -2$ are calculated as sum of the predicted sales for that period, the predicted sales of the period in which the goods are meant to arrive ($t = 0$), and the safety stock. From this sum, only the inventory level at the beginning of the regarded period needs to be subtracted. Thus, $O_{-2} = 51 + 50 + 634 - 685 = 50$. $I_{E_{-2}}$ results from subtracting 40 from 685.

The example shows a sharp increase in sales in period $t = 1$. This leads to a stockout at the end of that period. However, due to the application of the naïve forecasting method, the increase in sales in $t = 1$ also leads to very large order and to nearly 7,000 goods being delivered in period $t = 4$.

Table 2 Second toy example: Calculation of inventory level over time with *l* = 2, *s* = 634, forecasting method: Naive

t	D_t	I_{B_t}	y_t	$\hat{y}_t$	O_t	I_{E_t}
–2	51	685	40	51	50	645
–1	263	908	300	263	29	608
0	50	658	50	50	326	608
1	29	637	6,091	40	87	0
2	326	326	50	300	6,699	276
3	87	363	0	50	371	363
4	6,699	7,062	0	6,091	0	7,062
5	371	7,433	565	50	0	6,868

In the second step, the dynamic systems model calculates overstock and shortage costs based on the previously calculated inventory level development. Overstock costs C_{O_t} per period t are calculated by multiplying the difference of the average inventory level of that period and the safety stock with an overstock cost rate w, most likely the WACC. If the average inventory level is below the safety stock, which implies the lack overstock, the product becomes negative. Therefore, the product is set to be zero at minimal:

$$C_{O_t} = \max\left(\left(\frac{I_{E_t}+I_{B_t}}{2} - s\right) * w\ , 0\right). \tag{11}$$

The shortage costs C_{S_t} per period t are calculated by multiplying a shortage cost rate, e.g. the product's profit margin with the difference of the actual sales that could have taken place y_t and the inventory level at the beginning of that period I_{B_t}. Again, the product is limited to zero as there are no shortage costs, if the inventory level is higher than the sales:

$$C_{S_t} = \max\left(\left(y_t - I_{B_t}\right) * m, 0\right). \tag{12}$$

Finally, the overall costs C_t per period t are calculated as sum of overstock and shortage costs:

$$C_t = C_{O_t} + C_{S_t}. \tag{13}$$

Table 3 Continuation of the second toy example: Calculation of overstock and shortage costs, *s* = 634, *w* = 0.5% and *m* = 6%

t	I_{B_t}	y_t	I_{E_t}	C_{O_t}	C_{S_t}	C_t
–2	685	40	645	-	-	-
–1	908	300	608	-	-	-
0	658	50	608	0	0	0
1	637	6,091	0	0	327.24	327.24
2	326	50	276	0	0	0
3	363	0	363	0	0	0
4	7,062	0	7,062	32.14	0	32.14
5	7,433	565	6,868	32.58	0	32.58

Tab. 3 displays the continuation of the second toy example from Tab. 2. Based on the inventory levels at the beginning and end of each period, I_{B_t} and I_{E_t}, the overstock and shortage costs are calculated. The enormous increase in sales in period *t* = 1, which could have been observed in Tab. 2, leads to shortage costs of 327.24. The delivery of nearly 7,000 goods in period *t* = 4 however, leads to overstock costs in period *t* = 4 and the following periods.

4 EXPERIMENTAL SET-UP

The dynamic systems model is empirically tested using data from a German raw materials wholesaler. The company provided documentation of its sales from 1st September 2016 until 31st July 2022. Every product sale can be assigned to one of five material divisions. For each division, estimated profit margins were given, which serve as shortage costs. The WACC was used to measure the overstock costs. Because some products were sold very infrequently, filters were applied to the data. [31, 32] found that for time series forecasts on a monthly basis, at least 24 observations as training data are required in order to create reasonable forecasts. Rolling time series cross-validation was applied and for all time series, the test data were the sales from August 2021 until July 2022 (the last year of observations). The training data were accordingly the sales prior to the test data, the exact time period also depended on the lead time of the products.

Time series were included if they met the following criteria:

- Date of last sale was between May and July 2022 (to exclude products that are not sold anymore)
- Average sale frequency ≥ twice per month, meaning that the number of sales has to be higher than twice the number of months between the first and last sale (to exclude products with highly intermittent demand)
- The time series contained at least 36 observations plus the amount of lead time in observations (thus, at least 24 observations could be used for training the forecasting model).

These requirements limited the dataset to time series from 2,911 different products. Tab. 4 displays how many product time series from which division and with which lead time were included.

Table 4 Number of time series that met the inclusion criteria per product division and lead time

Division	Lead time in months				
	1	2	3	4	5
1	483	352	0	29	4
2	171	246	39	43	22
3	598	457	9	1	67
4	54	243	91	2	0
Sum	1,313	1,302	139	75	93

For these products, one-step rolling forecasts were created using the statistical forecasting methods naïve forecasting, ARIMA, and Holt-Winters' additive approach. They are considered classical sales forecasting methods and are easy to implement. As we propose a new dynamic systems model and focus on the comparison of forecast evaluation measures and not forecasting methods, we only consider statistical time series approaches. If forecasts happened to be negative, they were adjusted to zero. All forecasts were rounded to integers due to the application context. The forecasts were evaluated using the dynamic systems model and the statistical measures RMSE, MAE, and sMAPE. The evaluation with the dynamic systems model

was performed assuming three different service levels and thus three different safety stocks per product. Considered service levels were 90%, 95%, and 99%, corresponding to service factors Z equal to 1.3, 1.6, and 2.3 [30].

5 RESULTS

Each evaluation metric assessed one forecasting method to be most suitable for one product. Tab. 5 displays how often these assessments were coherent and how often different evaluation measures came to different results. The upper part of the table displays the assessments for all 2,911 time series. The middle and lower part show the results only for the divisions with either the lowest or highest profit margins.

Table 5 Percentage of time series for which the measures assessed the same forecasting method to be best

	Statistical evaluation			Economic evaluation with different service levels		
	RMSE	MAE	sMAPE	90%	95%	99%
For all-time series						
RMSE	100	75.37	59.64	55.82	55.14	54.14
MAE	75.37	100	67.06	55.62	55.93	56.51
sMAPE	59.64	67.06	100	43.15	42.39	41.39
90%	55.82	55.62	43.15	100	92.75	83.79
95%	55.14	55.93	42.39	92.75	100	89.42
99%	54.14	56.51	41.39	83.79	89.42	100
For the division with the lowest profit margin						
RMSE	100	73.51	59.31	55.09	54.51	53.93
MAE	73.51	100	67.37	57.20	58.15	58.35
sMAPE	59.31	67.37	100	43.19	43.95	42.99
90%	55.09	57.20	43.19	100	94.43	89.44
95%	54.51	58.15	43.95	94.43	100	92.90
99%	53.93	58.35	42.99	89.44	92.90	100
For the division with the highest profit margin						
RMSE	100	75.13	62.05	51.79	50.51	54.62
MAE	75.13	100	71.03	53.33	52.82	56.67
sMAPE	62.05	71.03	100	43.33	43.08	45.13
90%	51.79	53.33	43.33	100	91.54	77.44
95%	50.51	52.82	43.08	91.54	100	84.36
99%	54.62	56.67	45.13	77.44	84.36	100

The highest coherence among the statistical measures can be found between the RMSE and the MAE. This is plausible because both measures are absolute forecasting errors. However, even these two similar measures found different best methods for a quarter of the time series. The coherence among the dynamic systems model's evaluations with different safety stock is higher, up to 92.75%. This might imply that the safety stock level does not have a huge impact on the choice of the most suitable forecasting method. When comparing the dynamic systems model's evaluation with those of the statistical measures, it strikes that they rarely come to the same conclusions. The coherence between the RMSE and MAE and the dynamic systems model's evaluation ranges between 54% and 57%. However, for the sMAPE, it is even lower with values between 41% and 44%.

The lower part of Tab. 5 displays the coherence for the divisions with the lowest and the highest profit margin/ shortage costs. The comparison shows that the higher the profit margin, the more the results diverge. For the division with the lowest margin, there is a 92.90% coherence between the dynamic systems model with a service level of 95% and 99%. In the division with the highest profit margin, this value is 84.35%, considerably lower. Also the coherence with the statistical measures decreases with increasing profit margins.

Tab. 6 displays the average of the ranks the evaluation measures assigned to the forecasting methods. The lower the rank, the better the performance of the method.

Table 6 Average rank per forecasting method and performance metric

	Performance measures					
	Statistical evaluation			Economic evaluation with different service levels		
For all-time series						
Forecasting method	RMSE	MAE	sMAPE	90%	95%	99%
Naïve method	2.51	2.32	2.23	2.28	2.29	2.30
ARIMA	1.41	1.68	1.72	1.77	1.78	1.79
Holt-Winters	2.07	2.00	2.05	1.95	1.94	1.91
For l = 1						
Naïve method	2.49	2.27	2.18	2.24	2.24	2.25
ARIMA	1.40	1.71	1.73	1.75	1.77	1.81
Holt-Winters	2.11	2.02	2.09	2.00	1.99	1.94
For l = 5						
Naïve method	2.53	2.34	2.04	2.37	2.37	2.37
ARIMA	1.58	1.78	1.86	1.89	1.88	1.83
Holt-Winters	1.89	1.88	2.10	1.74	1.75	1.81

Among all evaluation measures and all-time series, ARIMA performed best. Interestingly, ARIMA is given a better evaluation by the statistical measures and a slightly worse evaluation by the dynamic systems models. The opposite happened for the Holt-Winters method, it was ranked better by the dynamic systems models than by the RMSE, MAE, and sMAPE. The naïve method is clearly the least accurate forecasting method. This is not surprising because it was used as a simple benchmark method.

The comparison of the ranks for different lead times shows that the naïve method performs better for a short lead time measured in all measures except the sMAPE. The change is intuitive because the naïve method assumes the next sales to be equal to the last observed sales. If the lead time is 1 and the demand quite stable, the naïve prediction can be reasonable. Moreover, it can be observed that the higher the lead time, the better performs Holt-Winters compared to ARIMA. For $l = 5$, the dynamic systems models rank Holt-Winters better than ARIMA. However, the statistical measures still rate ARIMA best method.

Fig. 4 displays the development of the average overstock and shortage costs per forecasting method with an increased service level. Naturally, the overstock costs increase with the service level and the shortage costs decrease. However, the decrease in shortage costs seems higher than the increase in overstock costs, especially going from a service level of 95% to 99%. Also striking is that the overstock costs are more than double the shortage costs. And the difference between the overstock costs of different forecasting methods is considerably higher than the difference in the shortage costs.

When comparing the forecasting methods, it is striking that the naïve method has both the highest overstock and shortage costs. ARIMA has the lowest overstock costs, but the second highest/ lowest shortage costs. Holt-Winters has subsequently the lowest shortage and the second highest/ lowest overstock costs.

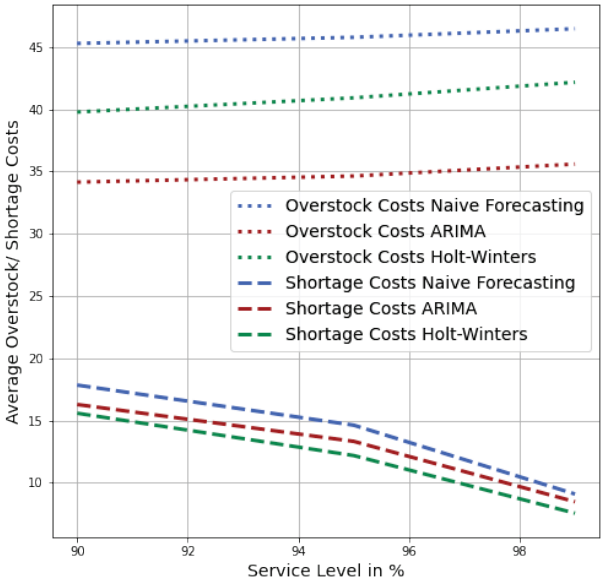

Figure 4 Overstock and Shortage Costs depending on the Service Level and Forecasting Method

6 DISCUSSION

The experiments showed a weakness of the traditional statistical evaluation measures. Even the commonly used RMSE, MAE, and sMAPE did not make coherent statements about the most suitable forecasting method for a time series. Moreover, for roughly half the time series, the cost-considering dynamic systems model came to another conclusion about the best method than the evaluations based upon statistical measures. However, considering costs is crucial for every business context.

The dynamic systems model describes the development of the inventory level over time and derives the resulting shortage and overstock costs. Nevertheless, the shortage costs might have been underestimated because in the real-world dataset, purchases that customers wanted to make but could not due to stockouts of the raw material wholesaler were not documented. Another weakness is that the model does not include order-related costs. It just assumes regular orders and fixed order costs. In practice, order costs vary and for example, quantity discounts could encourage ordering a larger number of goods at once.

Moreover, the assumptions stated in section 3 could be questioned. One assumption is e.g. that lead times are fixed. Recently during the pandemic, we have seen that lead times can significantly change and disturb a steady good supply.

7 CONCLUSION

Literature provides many different methods for evaluating a sales forecast. We have empirically shown on a real-world dataset of 2,911 time series that evaluations by the most commonly used measures RMSE, MAE and sMAPE (as an adaption of MAPE) differ considerably. Moreover, it has introduced a dynamic systems model to evaluate sales forecasts based on shortage and overstock costs. The costs are measured as the percentage of a product's price and can be adapted to the company's needs. In the experiments in this paper, the shortage costs were assumed to be equal to a product's profit margin and the overstock costs to a company's WACC. However, they could also be adapted to include penalty shortage costs for reputation loss in case of a stockout.

The main advantage of the dynamic systems model is that it considers the development of the inventory level over time. It can reproduce the costs that would arise from ordering based on a certain sales forecast. Thus, it helps to choose the most suitable forecasting method, minimizes inventory-related cost, maximizes profit and enables a smooth inventory flow.

In the experiments, three different service levels were regarded. It was shown that the evaluations of the forecasting methods do not differ considerably for the service levels. But, especially for products with a high profit margin, they have an impact on the evaluation. In future work, the service level could be integrated into the dynamic systems model as a tuning parameter to optimize the overall costs. Moreover, the dynamic systems model could be extended into a more complex model that considers variable lead times, storage, and order costs.

8 REFERENCES

[1] Schmid, L., Roidl, M., & Pauly, M. (2023). Comparing statistical and machine learning methods for time series forecasting in data-driven logistics--A simulation study. arXiv preprint arXiv:2303.07139.

[2] Gružauskas, V., Gimžauskienė, E., & Navickas, V. (2019). Forecasting accuracy influence on logistics clusters activities: The case of the food industry. *Journal of Cleaner Production, 240*(1), 118225. https://doi.org/10.1016/j.jclepro.2019.118225

[3] Wheelwright, S., Makridakis, S., & Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.

[4] Moosavi, J., Fathollahi-Fard, A. M., & Dulebenets, M. A. (2022). Supply chain disruption during the COVID-19 pandemic: Recognizing potential disruption management strategies. *International Journal of Disaster Risk Reduction*, 102983. https://doi.org/10.1016/j.ijdrr.2022.102983

[5] Petropoulos, F. et al., 2022. Forecasting: theory and practice. *International Journal of Forecasting, 38*(3), 705-871.

[6] Ickstadt, K. et al., (2022). Lernverfahren der Künstlichen Intelligenz zur Inwertsetzung von Daten: Auto-matisierte Erkennung und Prognose. In Silicon Economy: Wie digitale Plattformen industrielle Wertschöpfungsnetz-werke global verändern, 229-250. Berlin, Heidelberg: Springer Berlin Heidelberg. (in German) https://doi.org/10.1007/978-3-662-63956-6_11

[7] Huang, H., Pouls, M., Meyer, A., & Pauly, M. (2020). Travel time prediction using tree-based ensembles. In *Computational Logistics, the 11th International Conference, ICCL 2020*, Enschede, The Netherlands, September 28-30. Proceedings 11, 412-427. Springer International Publishing. https://doi.org/10.1007/978-3-030-59747-4_27

[8] Kiefer, D., & Ulmer, A. (2019). Application of Artificial Intelligence to optimize forecasting capability in procurement. In *Wissenschaftliche Vertiefungskonferenz: Informatik-Konferenz an der Hochschule Reutlingen*, 27. November 2019, 69-80.

[9] Ord, K., Fildes, R. A., & Kourentzes, N. (2017). *Principles of business forecasting*.
[10] Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages. *O. N. R. Memorandum No. 52*, Carnegie Institute of Technology, Pittsburgh USA.
[11] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science, 6*(3), 324-342. https://doi.org/10.1287/mnsc.6.3.324
[12] Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry: a review. *Mathematical Problems in Engineering*, 2013. https://doi.org/10.1155/2013/738675
[13] Jiang, H., Ruan, J., & Sun, J. (2021). Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast. In: *The 6th IEEE International Conference on Big Data Analytics (ICBDA)*, March, 5-8, Xiamen, China. https://doi.org/10.1109/ICBDA51983.2021.9403224
[14] Adam, S. P., Alexandropoulos, S. A. N., Pardalos, P. M., & Vrahatis, M. N. (2019). No free lunch theorem: A review. *Approximation and Optimization: Algorithms, Complexity and Applications*, 57-82. https://doi.org/10.1007/978-3-030-12767-1_5
[15] Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. E. (2013). A survey of forecast error measures. *World applied sciences journal, 24*(24), 171-176.
[16] Long, S. & Liu, Q. (2021). Research on New Energy Vehicle Sales Forecast and Product Optimization Based on Data Mining. In *the 2nd IEEE International Conference on Electronics, Communications and Information Technology (CECIT)*, 1019-1024. https://doi.org/10.1109/CECIT53797.2021.00181
[17] Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., & Zhang, R. (2020). ForeXGBoost: passenger car sales prediction based on XGBoost. *Distributed and Parallel Databases, 38*, 713-738. https://doi.org/10.1007/s10619-020-07294-y
[18] Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and computer-integrated manufacturing, 34*, 151-163. https://doi.org/10.1016/j.rcim.2014.12.015
[19] Li, J. (2022). A Feature Engineering Approach for Tree-based Machine Learning Sales Forecast, Optimized by a Genetic Algorithm Based Sales Feature Framework. In *the 5th IEEE International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 133-139. https://doi.org/10.1109/ICAIBD55127.2022.9820532
[20] Schmidt, A., Kabir, M. W. U., & Hoque, M. T. (2022). Machine learning based restaurant sales forecasting. *Machine Learning and Knowledge Extraction, 4*(1), 105-130. https://doi.org/10.3390/make4010006
[21] Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International journal of forecasting, 9*(4), 527-529. https://doi.org/10.1016/0169-2070(93)90079-3
[22] Dhakal, C. P. (2017). A naïve approach for comparing a forecast model. International Journal of Thesis Projects and Dissertations, 5(1), 1-3.
[23] Axsäter, S. (2015). *Inventory control* (Vol. 225). Springer. https://doi.org/10.1007/978-3-319-15729-0
[24] Waller, M. A. & Esper, T. L. (2014). *The definitive guide to inventory management: Principles and strategies for the efficient flow of inventory across the supply chain*. Pearson Education.
[25] Wu, J., Li, J., & Ou, H. Y. (2006, October). Impact of shortage cost in a risk-averse newsboy model. In *IEEE International Conference on Service Systems and Service Management, 1*, 278-282. https://doi.org/10.1109/ICSSSM.2006.320626
[26] Hasbullah, H., & Santoso, Y. (2020). Overstock Improvement by Combining Forecasting, EOQ, and ROP. *J. PASTI, 14*(3), 230-242. https://doi.org/10.22441/pasti.2020.v14i3.002
[27] Borade, A. B. & Bansod, S. V. (2011). Comparison of neural network-based forecasting methods using multi-criteria decision-making tools. *Supply Chain Forum: An International Journal, 12*(4), 4-14. https://doi.org/10.1080/16258312.2011.11517276
[28] Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research, 213*(2), 361-374. https://doi.org/10.1016/j.ejor.2010.11.024
[29] Fishwick, P. A. (Ed.). (2007). *Handbook of dynamic system modeling*. CRC Press. https://doi.org/10.1201/9781420010855
[30] Radasanu, A. C. (2016). Inventory management, service level and safety stock. *Journal of Public Administration, Finance and Law, 9*, 145-153.
[31] Lorek, K. S. & McKeown, J. C. (1978). The effect on predictive ability of reducing the number of observations on a time-series analysis of quarterly earnings data. *Journal of Accounting Research*, 204-214. https://doi.org/10.2307/2490418
[32] De Alba, E. & Mendoza, M. (2007). Bayesian forecasting methods for short time series. *The International Journal of Applied Forecasting, 8*, 41-44.

Authors' contacts:

Lara Kuhlmann, M.Sc.
(Corresponding author)
Graduate School of Logistics, Department of Mechanical Engineering and Department of Statistics, TU Dortmund University,
Leonhard-Euler-Straße 5,
44227 Dortmund, Germany
004915775241591, lara.kuhlmann@tu-dortmund.de

Markus Pauly, Prof. Dr.
Research Center Trustworthy Data Science and Cybersecurity and Department of Statistics, TU Dortmund University,
Otto-Hahn-Straße 14,
44227 Dortmund, Germany
pauly@statistik.tu-dortmund.de

APPENDIX

Acronyms of Error Measures

Acronym	Full name
MAE	Mean Absolute Error
MdAE	Median Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
MdAPE	Median Absolute Percentage Error
RMSPE	Root Mean Square Percentage Error
RMdSPE	Root Median Square Percentage Error
sMAPE	Symmetric Mean Absolute Percentage Error
sMdAPE	Symmetric Median Absolute Percentage Error
msMAPE	Modified Symmetric Mean Absolute Percentage Error
MRAE	Mean Relative Absolute Error
MdRAE	Median Relative Absolute Error
MASE	Mean Absolute Scaled Error
RMSSE	Root Mean Square Scaled Error
RMAE	Relative Mean Absolute Error
RRMSE	Relative Root Mean Square Error
LMR	Log Mean Squared Error Ratio
PB(MAE)	Percentage Better (MAE)
mRMSE	Normalized Root Mean Square Error
inRSE	Integral Normalized Mean Square Error
Std_AE	Standard Deviation of Absolute Error
Std_APE	Standard Deviation of Absolute Percentage Error

Derivation of Formula O_t

In Period *t*, there is no information about the actual demand (y_t) of future periods. Thus, it needs to be assumed that y_t equals the predicted demand ($\hat{y}_t$). For the periods after placing the order and before the goods arrive, we assume that the number of delivered goods equals the predicted demand.

Subsequently, we only need to consider the predicted demand for the current period *t* and the period *t+l* (lead time), in which the ordered goods will arrive. The sum of the inventory level at the beginning of period *t* (I_{B_t}) and the ordered goods (O_t) should equal the sum of the demand for the current period ($\hat{y}_t$), the demand for *t+l* ($\hat{y}_{t+l}$) and the safety stock (*s)*:

$$I_{B_t} + O_t = \hat{y}_t + \hat{y}_{t+l} + s. \tag{14}$$

Rearranging the formula and considering that the number of ordered goods cannot be negative leads to:

$$O_t = \max\left(\hat{y}_{t+l} + s + \hat{y}_t - I_{B_t}, 0\right). \tag{15}$$