Optimising forecasting models for inventory planning

Abstract

Inaccurate forecasts can be costly for company operations, in terms of stock-outs and lost sales, or over-stocking, while not meeting service level targets. The forecasting literature, often disjoint from the needs of the forecast users, has focused on providing optimal models in terms of likelihood and various accuracy metrics. However, there is evidence that this does not always lead to better inventory performance, as often the translation between forecast errors and inventory results is not linear. In this study, we consider an approach to parametrising forecasting models by directly considering appropriate inventory metrics and the current inventory policy. We propose a way to combine the competing multiple inventory objectives, i.e. meeting demand, while eliminating excessive stock, and use the resulting cost function to identify inventory optimal parameters for forecasting models. We evaluate the proposed parametrisation against established alternatives and demonstrate its performance on real data. Furthermore, we explore the connection between forecast accuracy and inventory performance and discuss the extent to which the former is an appropriate proxy of the latter.

Keywords: Forecasting, inventory management, optimisation, likelihood, simulation

1. Introduction

In managing inventory it is critical to quantify the demand uncertainty, typically encapsulated in forecasts of the demand over the lead time period. There is an intuitive link between accurate forecasting and meeting the demand. When the demand is unusually high, compared to the expected, i.e. the forecast, it may result in a stockout, while when the demand is below expectations, additional holding costs can arise. Nonetheless, whether a deviation of demand from the expectation leads to a stockout or excess inventory holding costs is conditional on the current stock on hand. Given sufficient stock, a stockout can be averted, or equivalently high holding costs.

In a realistic setting, where the demand process is unknown, we rely on forecasts to generate approximations of the expected realised demand over the lead time. However, the generation of forecasts is often rather disjoint from the needs of the forecast users. Forecasting methods, such as the exponential smoothing method, are typically optimised on the in-sample mean squared errors (MSE), which when used as a cost function for the parameter optimisation leads to optimal forecasts for the mean of the demand process (Gneiting, 2011). Similarly, when a model is present, we typically use maximum likelihood estimation, which is again based on quadratic errors, and therefore results in optimal forecasts for the mean of the demand process. This ensures unbiased in-sample forecasts, however this does not guarantee good out-of-sample behaviour (Barrow and Kourentzes, 2016), i.e. accurate or unbiased forecasts. Alternative costs have been explored, such as the mean absolute error (MAE, Gardner Jr, 2006), which are more robust against outlying demand events. Absolute errors result in optimal forecasts for the median of the demand process (Gneiting, 2011). However, forecasts for inventory management need to be neither solely optimal on the mean nor median of the demand process.

In fact, there are two elements where classical cost functions fail in an inventory management setting. First, they satisfy their conditions in-sample; and second, inventory decisions are conditional on the inventory position and the ordering policy, neither of which are accounted by MSE, MAE or similar. To exemplify the first point, we can consider the following thought experiment. Consider a simple linear regression, for which we can always obtain the optimal parameters analytically, when the cost is quadratic. Suppose we fit the model on n data points. In the next forecasting cycle we can refit the model using $n + 1$ data points, to consider the most recent demand information. The two estimated set of parameters are bound to be different. The degree of difference depends on how close the model is to the underlying data generating process, which for any realistic case is unknown and typically more complex than the forecasting models we use to approximate it. This is a well documented problem and has led to extensive research in shrinkage estimators, such as lasso regression (Tibshirani, 1996). More recently Kourentzes and Trapero (2018) showed that multi-step ahead quadratic cost functions are a form of univariate shrinkage, which will also mitigate the sampling effects on parameter estimation. Nonetheless, these approaches do not address the second limitation discussed above.

In this paper we propose optimising the forecasts directly on the decision variable. Therefore, instead of relying on some metric of divergence between the fitted values of the forecast functions and the observed historical demand, we conduct an inventory simulation and optimise the forecast function parameters, so as to maximise the relevant inventory performance metrics. This follows the ideas of Simulation Optimisation (Amaran et al., 2016), which lends itself well when the problem at hand is stochastic and there is uncertainty or very high complexity of the error surface. Following the proposed approach, we minimise the risk of not satisfying incoming demand, given the inventory policy in place and accounting for the inventory position. Naturally, although the proposed approach aims to address the disconnect between the forecast and inventory management objective, it introduces new complications in the modelling process, which we discuss in this paper.

We evaluate the proposed optimisation on a real case with data from a UK manufacturer of fast moving consumer goods, and demonstrate that the proposed approach results in superior inventory performance, but inferior forecast accuracy compared to conventional cost functions. Given that conventional costs are designed to maximise forecast accuracy, particularly when the cost is of the same order, for example quadratic, this is not unreasonable. This is also in agreement with observation in the literature, that the most accurate forecast may not always have the best inventory performance (for example see, Kourentzes, 2013, 2014, and references therein). Furthermore, the proposed approach has the advantage that it offers a transparent connection between the parametrisation of the forecasts and their use, which we argue is appealing for practice.

The rest of the paper is organised as follows. Section 2 provides a summary of the background research on the connection between forecast accuracy, estimators and inventory performance. Section 3 outlines the proposed approach, followed by section 4 that presents the experimental design and the results. Section 5 expands on the results, providing some insights into the implications of using the proposed estimator, followed by concluding remarks in section 6.

2. Background research

Although evaluation is paramount in the forecasting research, often this is restricted to looking at the forecast accuracy, which is assumed to be a reasonable proxy for the decisions that forecasts support (Ord et al., 2017). Nonetheless, in the literature, there are papers that have questioned this assumption, and arguing for directly using stock control metrics in the context of inventory management (for examples see, Gardner, 1990a; Syntetos and Boylan, 2005, 2006; Teunter and Duncan, 2009; Syntetos et al., 2010; Kourentzes, 2013; Syntetos et al., 2015a; Sagaert et al., 2018).

Substantial work has been done in the intermittent demand literature, where forecast evaluation is problematic when using conventional error metrics (Kolassa, 2016). Kourentzes (2013) evaluates the use of neural networks for intermittent demand forecasting and finds that although they are inferior in terms of forecast accuracy, ranking almost last amongst the benchmarks considered, the opposite result emerges when looking at inventory performance metrics. Kourentzes (2014) similarly finds that seemingly different performing methods in terms of accuracy, result in insubstantial differences when evaluated in inventory performance terms. These findings are in agreement with aforementioned references in the intermittent demand literature, raising the question whether forecasts optimal on typical cost functions add considerable value to the subsequent decisions they support, and in particular inventory decisions.

Sanders and Graman (2009) explores the connection between forecasting performance and the resulting decisions in depth. An interesting aspect of this work is that they consider both the accuracy and the bias aspect of forecasting performance. They find that bias has substantially larger impact on the performance of the subsequent decisions compared to accuracy.

It is useful to consider the effect of minimising MSE when generating forecasts. Quadratic loss makes forecasts optimal for the mean of the demand distribution (Gneiting, 2011). At the same time, the bias-variance decomposition indicates that minimising MSE results in minimising the variance of the forecasts, as well as the in-sample squared bias (defined by the mean error), resulting in in-sample unbiased predictors (Friedman et al., 2001). However, anticipating the out-of-sample performance of MSE optimal forecasts is far from straightforward. For example, Barrow and Kourentzes (2016) provide evidence of how weak is the connection between in-sample and out-of-sample forecast error, where their distributions can differ considerably. They proceed to suggest that forecast combination is a potential remedy. Forecast combination can be perceived as shrinkage operator (Elliott et al., 2013; Kourentzes et al., 2018). This connects to the bias-variance trade-off, where the modeller has the challenge of specifying a forecast that balances in-sample over- and under-fitting, given that the true data generating process is unknown, so as to achieve good out-of-sample performance (Friedman et al., 2001). Very accurate in-sample forecasts, and by construction in-sample unbiased if they are optimal on MSE, may be over-fit, which results in poor out-of-sample performance. Optimising forecast parameters on MSE, or similar cost functions, does not provide with any warnings of over-fitting, and it is left up to the modeller to use a variety of model building diagnostics to detect the problem. Therefore, we argue that the challenge of producing well performing forecasts in the out-of-sample, particularly in terms of bias, largely falls on the way we parametrise forecasts. Shrinkage approaches have gained popularity, as they connect model specification and estimation in a single step, and make the bias-variance trade-off evident to the modeller (Tibshirani, 1996; Friedman et al., 2001). This motivates our proposed approach for alternative forecast parametrisation, outlined in the next section.

3. Optimising directly on inventory performance

A high level overview of the proposed approach for estimating model parameters is provided in the flowchart in figure 1. This resembles conventional derivative free optimisation, but instead of minimising some historical demand fitting error, at each iteration an inventory simulation is performed, tracking some appropriate performance indicator, which is then used to evaluate the cost of the optimisation and inform further updating of the forecasting model parameters or stopping the optimisation process.

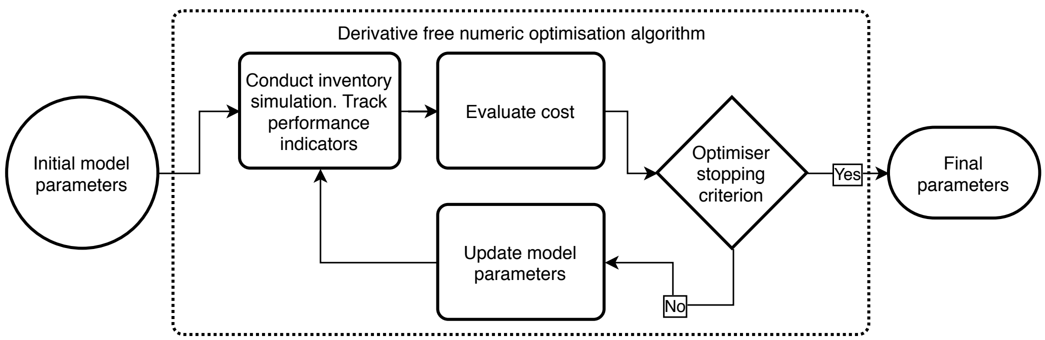

Figure 1: Flowchart of the proposed parameter estimation approach.

It is obvious that the modeller needs to provide the various details for the simulation, such as the relevant inventory policy, lead times, and so on, which are specified so as to match the realistic inventory decision that is faced by the company. In contrast to typical simulation optimisation, in this case there is minimal need for calibration, as inventory management is a well studied and understood process. Typically the main sources of uncertainty are related to describing the demand process and its uncertainty, which are the forecasting related questions that we are trying to optimise in the first place. Other elements of the simulation are provided by the business context, and depending on the complexity of the business environment and the desired degree of realism, this can often be captured with high certainty.

It is useful to contrast the sample requirements for conventional time series modelling optimisation and the proposed one, as well as, what this implies for the optimal parameters. When estimating the parameters of a forecasting model or method, conventionally, we consider some summary performance statistic, such as the MSE, which becomes the cost function for the optimiser. The MSE is the sample fitting error variance, which estimates the unobserved error variance that we want to minimise, and being a random variable we require adequate sample to obtain a good estimate. At the same time, since the cost function is minimised when the differences between the observed and the predicted values are minimised, this becomes an approximation exercise. As the parameters of the forecast function increase, then it can approximate more flexibly the sample observations. If it correctly describes the underlying data generating process then, since the parameters are random variables themselves, additional sample increases our confidence in the estimates. If the forecast function has omitted terms, then there are limits to the quality of the approximation, which will become more evident as the sample size increases. Alternatively, if there are additional terms, then the forecast function will over-fit, however this effect will lessen as sample size increases, with the unnecessary parameters eventually converging to zero. In reality, the underlying process is always unknown, and our forecast functions may have simultaneously omitted and superfluous terms.

Switching to the proposed simulation based optimisation, we still summarise the performance using a performance metric, which itself is a random variable. From that standpoint, additional sample will increase of confidence in the estimate, similar to the conventional case. However, the cost function no longer prescribes an approximation exercise to the observed data and, therefore, there is no direct connection between the model terms and the sample size. Indirectly, a reasonable approximation of the data generating process is still required, but one has to note that this is less straightforward than in the conventional case. For example, consider operational production planning, where a consistent forecast across time may be more beneficial than a very accurate, but volatile, forecast (Sagaert et al., 2018; Fildes and Kingsman, 2011). Nonetheless, if the quality of the forecast is very poor, then that impacts negatively on the operations. The exact transformation of the forecasts to the decision can be rather complex, or even unknown, depending on the decision. This is precisely when simulation optimisation is beneficial (Amaran et al., 2016).

Specifically for the inventory management case, as the forecasts and their variance are used to inform re-ordering decisions and the amount of safety stock, reducing the bias of the forecast becomes more important to accuracy. A forecast that is very accurate in-sample, but gives biased out-of-sample forecasts can severely impact negatively the inventory performance. On the other hand, a less accurate forecast that remains fairly unbiased in the out-of-sample can be preferable, even if the forecast function has omitted terms. Similarly, this approach, as it is not minimising the approximation error, will have less tendency to over-fit, even when unnecessary terms are included in the forecast function.

A final important consideration is what should be the appropriate inventory performance metric. Ideally, one should use the overage and underage costs directly, however these are not always easy to obtain, particularly the underage cost. Therefore, as is the case inventory performance evaluation, we can consider associated metrics, such as the cycle service level and the fill rate. The cost in that case becomes the difference between the target and realised for the sample period. Both the service level and the fill rate blend the overage and the underage cost into a single metric, which effectively reduces a multi-objective optimisation to a conventional single objective one.

In the following section we describe the implementation details of the optimisation, and evaluate the proposed parametrisation against established benchmark approaches, to demonstrate the merit of the idea.

4. Empirical evaluation

4.1. Case study dataset

To evaluate the performance of the proposed parametrisation we use a real case study. Our data originate from a UK manufacturing firm that specialises in household cleaning and personal hygiene products. The company serves retailers in the European market and has production facilities distributed in the continent. They operate on a weekly inventory planning cycle and their lead time is typically between 3 to 5 weeks.

Our dataset contains 229 items with 173 weekly sales each. An exploration of the sales data reveals not strong evidence of seasonality or trends. We retain the last $m = 52$ weeks as a test set, to evaluate the resulting forecasting and inventory performance of the evaluated model parametrisation alternatives.

4.2. Forecasting setup

Given the structure of the time series, we use local level exponential smoothing, that is the state space model form of the popular single exponential smoothing method (Hyndman et al., 2002):

$$\hat{y}_{t+h} = l_t, \quad \text{and} \quad h \geq 1, \tag{1}$$

$$l_t = l_{t-1} + \alpha\varepsilon_t, \tag{2}$$

where $\hat{y}_{t+h}$ is the h-step ahead forecast from period t, l_t the estimate of the local level, α the smoothing parameter and $\varepsilon_t \sim N(0, \sigma)$. Observe that the forecast $\hat{y}_{t+h}$ is conditional on the observed sample up to period t, the smoothing parameter α and the initial level l_0, which corresponds to the value of l_{t-1} when $t = 1$. The conditional demand over the lead time is simply $\hat{Y}_{t+L} = \sum_{i=1}^{L} \hat{y}_{t+i}$, updated at every period. From the state space framework, we can construct the expression for the conditional variance over the lead time, L, given parameter α (Hyndman et al., 2008, p. 92):

$$V_L = \sigma^2 L \left[1 + \alpha(L-1) + \frac{1}{6}\alpha^2(L-1)(2L-1)\right]. \tag{3}$$

Here, L is constant and known, and any review time is included in it. Note that equation (3) does not account for the uncertainty due to the estimation of α. Prak et al. (2017) and Prak and Teunter (2018) show that for small sample size this can lead to high inventory costs, as the total uncertainty is underestimated. Whereas, when there is adequate estimation sample, the parameter uncertainty becomes sufficiently small to have minuscule impact.

We consider a number of alternatives to estimate the smoothing parameter α and the initial level l_0. Given the state space formulation we can derive the likelihood and maximise it to obtain the optimal parameters given the available sample. This turns out to be equivalent to the well known Mean Squared Error, given the additive nature of the model innovations (Hyndman et al., 2008, p. 69):

$$\text{MSE} = \frac{1}{n-1}\sum_{t=1}^{n-1}\left(y_{t+1} - \hat{y}_{t+1}\right)^2, \tag{4}$$

where n is sample size of the training set and y_t the observed demand at period t. Kourentzes and Trapero (2018) have argued, echoing similar views by Xia et al. (2011) and Chatfield (2000), that maximising the likelihood is meaningful given the assumption that the underlying model is true for the demand process. This is often adequate for practical considerations. However, when this assumption is strongly violated, minimising (4) will result in model parameters that are adequate only for short term predictions. Kourentzes and Trapero (2018) proceed to show that using multi-step cost functions can result in increased forecast accuracy (also see empirical evidence by Clements and Hendry, 1998; Pesaran et al., 2011), and results in a particular form of univariate parameter shrinkage. This also mitigates the parameter estimation uncertainty. One can minimise the h-steps ahead forecast error, or the average of 1 to h-steps ahead error, with the latter enforcing more aggressive shrinkage:

$$\mathrm{MSE}_{t+h} = \frac{1}{n-h} \sum_{t=1}^{n-h} \left(y_{t+h} - \hat{y}_{t+h}\right)^2, \tag{5}$$

$$\mathrm{MSE}_{t+1 - t+h} = \frac{1}{h} \sum_{i=1}^{h} MSE_{t+i}, \tag{6}$$

Finally, considering that the requirement in inventory management is accurate forecasts over the lead time, one can minimise these directly:

$$\mathrm{MSE}_{t+L} = \frac{1}{n-L} \sum_{t=1}^{n-L} \left(Y_{t+L} - \hat{Y}_{t+L}\right)^2, \tag{7}$$

where actual demand over lead time $Y_{t+L} = \sum_{i=1}^{L} y_{t+i}$.

Finally, the last alternative we consider is to use the inventory simulation approach we propose, where the cost function to be minimised is simply:

$$\text{Cost} = (p - \hat{p})^2, \tag{8}$$

where p is the target cycle service level and $\hat{p}$ is the realised one, for the in-sample period, based on an inventory simulation for given model parameters.

Given the model expression for the lead time variance (3), we need an estimate for the σ, which given parameters α and l_0 is:

$$\hat{\sigma} = \frac{1}{n-2} \sum_{t=1}^{n} (y_{t+1} - \hat{y}_{t+1})^2. \tag{9}$$

We divide by $n - 2$ to account properly for the degrees of freedom.

We consider these four estimation approaches, alongside the proposed one, to rigorously benchmark the performance of the inventory simulation optimisation approach. Furthermore, this study provides the opportunity to evaluate the impact of the alternative estimation approaches to the inventory, which has not been done in the literature.

4.3. Evaluation setup

We consider lead times of 3 and 5 weeks and perform a rolling origin evaluation. At every period we optimise the model parameters and produce the relevant forecasts. Once this is done, we increase the sample by one observation and repeat. This is repeated until all the test set is used. We consider four forecasting performance metrics, two measuring the magnitude of forecast errors and two the magnitude of bias. For each case, one metric is recording the errors per period, and the other the cumulative errors over lead time.

$$\text{RMSE}_h = \frac{1}{m-h+1} \sum_{t=n+1}^{n+m-h} \sqrt{\frac{1}{h} \sum_{i=1}^{h} \left(y_{t+i} - \hat{y}_{t+i}\right)^2}, \tag{10}$$

$$\text{RMSCE}_h = \frac{1}{m-h+1} \sum_{t=n+1}^{n+m-h} \sqrt{\left(Y_{t+h} - \hat{Y}_{t+h}\right)^2}, \tag{11}$$

$$\text{AME}_h = \left| \frac{1}{m-h+1} \sum_{t=n+1}^{n+m-h} \left(\frac{1}{h} \sum_{i=1}^{h} \left(y_{t+i} - \hat{y}_{t+i}\right) \right) \right|, \tag{12}$$

$$\text{AMCE}_h = \left| \frac{1}{m-h+1} \sum_{t=n+1}^{n+m-h} \left(Y_{t+h} - \hat{Y}_{t+h} \right) \right|. \tag{13}$$

RMSE and AME are calculated on the trace forecast errors from $t+1$ to $t+h$-steps ahead, for each $m-h+1$ forecast origins, where m is the sample size of the test set. Likewise, RMSCE and AMCE are calculated on the difference of the cumulative actuals and forecasts over the lead time, for each forecast origin. Observe that AME and AMCE are measuring the same quantity, as the summations of y_{t+i} and $\hat{y}_{t+i}$ in (12) can be rearranged to give (13). For this reason we only retain AME in the discussion. These errors are scale-dependent. We transform them to scale independent by dividing for each time series the resulting metric for estimation method A with a benchmark method B, for which we use the standard 1-step ahead in-sample MSE. These ratios are then summarised across the q time series of the dataset using the geometric mean:

$$\text{Rel}M_h = \left(\prod_{i=1}^{q} \frac{M_h^A}{M_h^B} \right)^{\frac{1}{q}}, \tag{14}$$

where $M_h = (\text{RMSE}_h, \text{RMSCE}_h, \text{AME}_h)$. When the value of the relative metric is below one the method A improves over the benchmark by $(1 - \text{Rel}M_h)100\%$.

Beyond the forecast performance metrics, we consider the inventory performance of the alternative estimation methods by constructing an inventory simulation. We impose an order-up-to policy and measure the outcome for $p = (90\%, 95\%, 99\%)$ service levels, matching targets in the case study company. The order-up-to level S is calculated as:

$$S = \hat{Y}_{t+L} + \Phi^{-1}(p)\sqrt{V_L}, \tag{15}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the cumulative normal distribution and V_L is calculated as in (3).

We keep track of the realised service level, the stock-on-hand and the stock-out, scaled by the observed in-sample mean demand, to make these metrics scale independent. We then average these across all time series in the dataset. Any unmet demand is considered lost, which is a reasonable assumption for the products of the case company.

4.4. Forecasting performance results

First, we report the forecasting performance of the competing estimation approaches. Table 1 summarises the various forecasting performance metrics. These are provided at an aggregate level across all time series, for the two different forecast horizons. Note that there are three results for the estimation based on the inventory simulation, one for each target service level. For each horizon and each metric, the best performing estimation method is highlighted in boldface.

Table 1: Forecasting performance metrics

Cost function	Accuracy		Bias
	RelRMSE	RelRMSCE	RelAME
		Horizon 3	
MSE	1.000	1.000	1.000
MSE_{t+h}	0.995	0.990	1.016
$MSE_{t+1—t+h}$	**0.994**	**0.987**	1.060
MSE_{t+L}	0.995	0.990	1.067
Inventory (90%)	1.033	1.086	0.678
Inventory (95%)	1.033	1.086	0.524
Inventory (99%)	1.062	1.157	**0.376**
		Horizon 5	
MSE	1.000	1.000	1.000
MSE_{t+h}	0.995	0.987	0.968
$MSE_{t+1—t+h}$	**0.991**	**0.974**	1.043
MSE_{t+L}	0.994	0.980	1.108
Inventory (90%)	1.009	1.028	0.745
Inventory (95%)	1.008	1.031	0.632
Inventory (99%)	1.023	1.089	**0.473**

The results across the two horizons are qualitatively similar. In terms of accuracy, irrespective if we focus at trace (RelRMSE) or cumulative (Rel-

RMSCE) errors over the horizon, $\text{MSE}_{t+1-t+h}$ performs best, although only marginally compared to the other MSE approaches. When we look at the bias metrics, again for both cases, the inventory based estimation substantially outpeforms all MSE based alternatives. The reduction of bias from the conventional MSE ranges between 25% to 60%. This is achieved with only up to 9% reduction in accuracy. We anticipate this reduction in bias to have a significant impact on the inventory performance.

Figure 2 provides violin plots of the three forecasting metrics, across time series, for $h = 3$. The geometric mean performance for each estimation method is indicated by a horizontal bar. The distributions have been winsorized at 5% to improve the visual clarity, which also explains the increased concentration of errors at the lower and higher values of the plots. We do not provide the plot for $h = 5$, as it is similar qualitatively and does not provide any additional insights. Also note that we provide a single set of results for the inventory based estimator, by calculating the geometric mean across the results for the different target service levels. We do that to avoid cluttering the figure, while providing minimal additional insights. Inspecting the figure provides a clear view that the differences between the MSE variants are not significant. On the other hand, the inventory based parametrisation provides significantly different results across all three metrics.

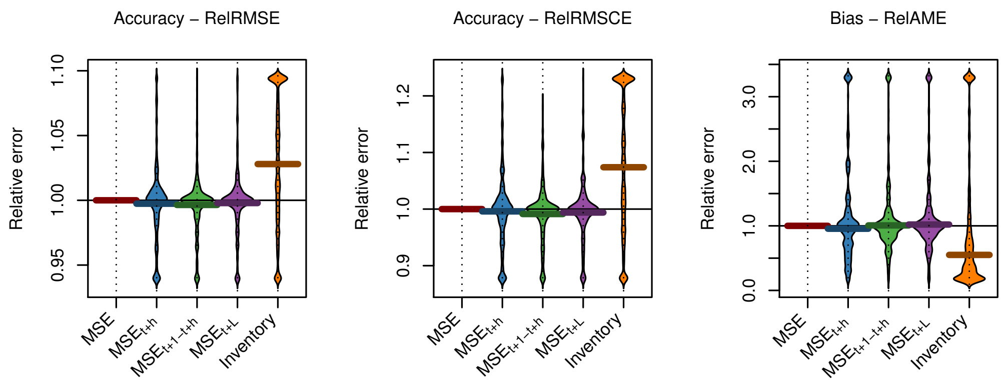

Figure 2: Violin plots of forecasting performance across time series, for $h = 3$.

4.5. Inventory performance results

We proceed to explore the performance in terms of inventory by means of trade-off curves (Gardner, 1990b; Syntetos et al., 2015b), as provided in figures 3 and 4. The top-left subplot provides the scaled inventory on hand against the service level deviation. We use the deviation to better highlight cases of under- and over-coverage. Ideally, any curves should have zero deviation and as little inventory as possible. Similarly, the top-right plot provides the scaled out of stock against the service level deviation. The bottom-left plot provides the trade-off curve between scaled inventory on hand and out of stock volume. In this subplot, the ideal solution would be having zero out of stock, while holding minimal inventory, and therefore any curves closer to the origin $(0, 0)$ dominate others.

Both figures 3 and 4 provide similar insights. The proposed inventory simulation based parameter estimation results in the smallest deviation from

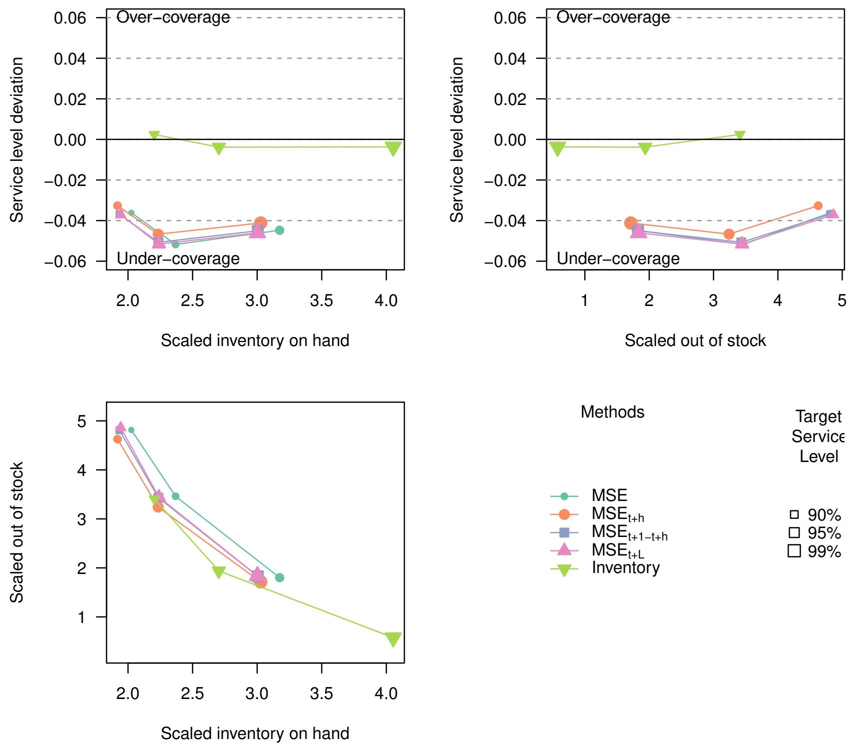

Figure 3: Inventory performance plots for $h = 3$.

the target service level, with only limited increase of the stock on hand and reduction of out of stock. This results in dominating trade-off curves over all other estimators. Considering the benchmarks, note that the conventional MSE based optimisation performs the worst, albeit with small differences from the remaining MSE variants. The superiority of the proposed parametrisation is stronger for $h = 5$.

The inventory performance results follow the discussion of the forecasting performance results, where the substantial gains in terms of out-of-sample bias, with only minimal degradation of accuracy, are translated in superior

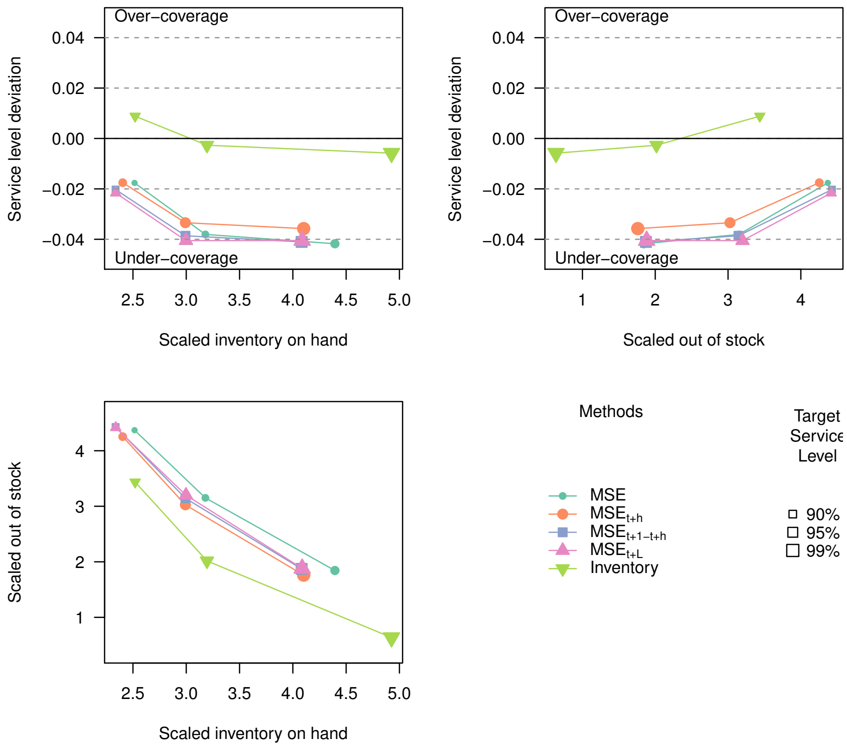

Figure 4: Inventory performance plots for $h = 5$.

inventory performance as expected.

5. Discussion

An advantage of MSE-based cost functions is that they typically result in well behaved convex error surfaces. In particular for the local level exponential smoothing model they prescribe a globally optimal set of α and l_0 values, for given historical demand.

It is useful to explore the resulting error surface of the proposed optimisation. We simulate the error surfaces for a time series of our case study, for both MSE and the proposed cost, plotted in figure 5 using contours. The resulting optimal set of parameters are provided as well. Observe how different the two optimal solutions are.

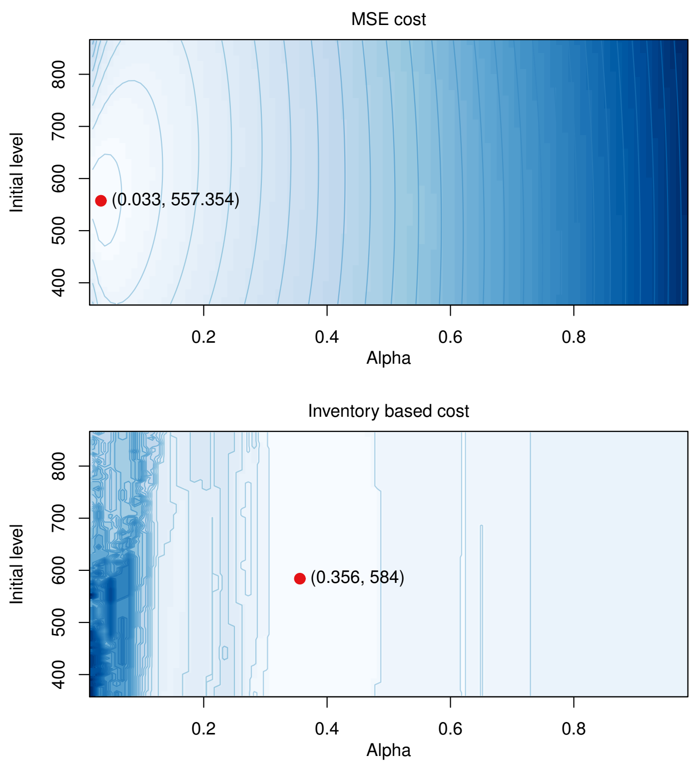

Figure 5: MSE and inventory based error surfaces, together with optimal parameters. Lighter colour denotes lower errors.

Observe that although the MSE error surface is convex and smooth, with a global minimum, this is not the case for the other cost. The inventory based error surface is more complex, non-monotonic and has multiple local minima. This makes the initial starting point of the optimiser important, as poorly chosen sets of starting values can result in low quality solutions.

Figure 6 exemplifies this by providing the error surfaces for two different time series, together with multiple starting points for the optimisation and associated end points, which are local minima. In the top subplot we observe that most converge around the same solution. However, in the second case, the large plateaus in the error surface result in many diverse local optima. In fact, many of the poor initialisation points at the right side of the subplot, associated with large α values, are not able to escape the plateau at all.

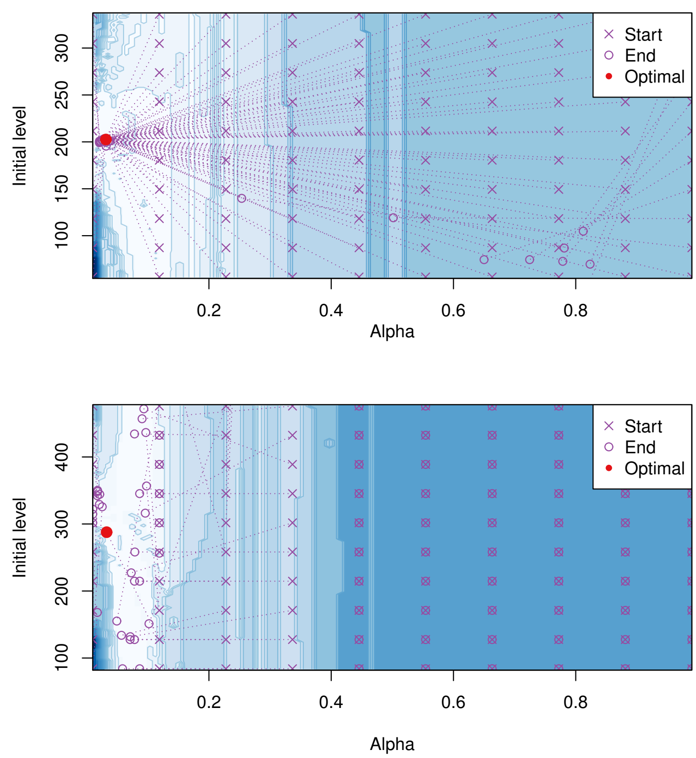

Figure 6: Inventory based error surfaces for two different time series. Multiple optimisation starting points are indicated, together with the corresponding local minima. Lighter colour denotes lower errors.

Our approach to avoid these issues was to initialise the optimisation multiple times, as indicated in figure 6 that forces a thorough search of the solution space.

Beyond the specifics of the error surface, using an inventory simulation based optimisation has further implications for the forecasting modelling methodology. Although it is applicable to both model and method based forecasts, using models is attractive as they provide analytical variance expressions that are crucial for the calculation of safety stocks in an inventory setting. Moreover, using model families is particularly helpful for selecting the appropriate model specification, which is typically done by using information critieria, such as the Akaike Information Criterion (Burnham and Anderson, 2003). However, these require that the likelihood of the model is maximised. This is no longer the case, which makes model selection and specification using information criteria invalid. This limitation is true for the various other cost functions described in section 4.2, and more widely in the literature, as well. There are various ways one could overcome this, also applicable when we rely on method based forecasts, such as using cross-validated errors, model combination and selection heuristics (for a discussion and comparison of common approaches see Kourentzes et al., 2018), potentially enhanced by domain knowledge (Petropoulos et al., 2018). How to best perform model selection and combination is an open question that requires additional research, with implications for the wider literature.

6. Conclusions

In this paper we proposed to parametrise forecasting models using a cost function inferred directly from the inventory decisions, instead of minimising the fitting error on past historical demand, as is the norm. We found that this resulted in lower forecast accuracy than conventional quadratic loss functions, which is not surprising. The loss in terms of accuracy was as high as 9%. However, there were substantial gains in terms of out-of-sample forecast bias, of up to 62% improvement. When translated to inventory performance, the proposed cost function achieved minimal differences between the target and realised service levels, in contrast to the benchmarks. Furthermore, the resulting out of stock-inventory on hand trade-off curves dominated other alternatives.

The inventory gains come at a cost of a more difficult optimisation problem. We demonstrate that the resulting error surfaces have multiple local minima and large plateaus, and suggested using multiple starting points for the optimiser to achieve an effective search of the solution space. It is of interest to explore alternative optimisation strategies that may result in more efficient parameter space search.

In this case study, given the available data, there was no need to complicate the modelling process by considering alternative exponential smoothing model forms. Nonetheless, the proposed optimisation makes the use of information criteria invalid and one has to revert to more generally applicable strategies, such as using cross-validated routines. Future research should explore whether the proposed approach can facilitate novel model selection strategies, driven directly by the relevant decision being simulated.

Finally, here we consider the inventory decision to construct the appropriate simulation based cost function. An attractive property of this modelling question is that inventory decisions are high frequency, and therefore occur multiple times within the in-sample period, thus providing adequate sensitivity for the optimisation. The proposed approach allows exchanging the inventory decision with other alternatives, and therefore providing a general framework for how to parametrise forecasts directly on the decisions, as long as an associated simulation can be constructed. Exploring further the generality of the approach for parametrising forecasts in other decision making contexts will be valuable.

Acknowledgment

This work was partially supported by the European Regional Development Fund and Spanish Government (MINECO/FEDER, UE) under the project with reference DPI2015-64133-R.

References

Amaran, S., Sahinidis, N. V., Sharda, B., Bury, S. J., 2016. Simulation optimization: a review of algorithms and applications. Annals of Operations Research 240 (1), 351–380.

Barrow, D. K., Kourentzes, N., 2016. Distributions of forecasting errors of forecast combinations: implications for inventory management. International Journal of Production Economics 177, 24–33.

Burnham, K. P., Anderson, D. R., 2003. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.

Chatfield, C., 2000. Time-series forecasting. Chapman and Hall/CRC.

Clements, M., Hendry, D., 1998. Forecasting economic time series. Cambridge University Press.

Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. Journal of Econometrics 177 (2), 357–373.

Fildes, R., Kingsman, B., 2011. Incorporating demand uncertainty and forecast error in supply chain planning models. Journal of the Operational Research Society 62 (3), 483–500.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The elements of statistical learning, 10th Edition. Vol. 1. Springer series in statistics New York, NY, USA:.

Gardner, E. S., 1990a. Evaluating forecast performance in an inventory control system. Management Science 36 (4), 490–499.

Gardner, Jr., E. S., 1990b. Evaluating forecast performance in an inventory control system. Management Science 36 (4), 490 – 499.

Gardner Jr, E. S., 2006. Exponential smoothing: The state of the artpart ii. International journal of forecasting 22 (4), 637–666.

Gneiting, T., 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106 (494), 746–762.

Hyndman, R., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of forecasting 18 (3), 439–454.

Kolassa, S., 2016. Evaluating predictive count data distributions in retail sales forecasting. International Journal of Forecasting 32 (3), 788–803.

Kourentzes, N., 2013. Intermittent demand forecasts with neural networks. International Journal of Production Economics 143 (1), 198–206.

Kourentzes, N., 2014. On intermittent demand model optimisation and selection. International Journal of Production Economics 156, 180–190.

Kourentzes, N., Barrow, D., Petropoulos, F., 2018. Another look at forecast selection and combination: evidence from forecast pooling. International Journal of Production Economics.

Kourentzes, N., Trapero, J. R., 2018. On the use of multi-step cost functions for generating forecasts. Department of Management Science Working Paper Series.

Ord, J. K., Fildes, R., Kourentzes, N., 2017. Principles of Business Forecasting, 2nd Edition. Wessex Press Publishing Co.

Pesaran, M. H., Pick, A., Timmermann, A., 2011. Variable selection, estimation and inference for multi-period forecasting problems. Journal of Econometrics 164 (1), 173–187.

Petropoulos, F., Kourentzes, N., Nikolopoulos, K., Siemsen, E., 2018. Judgmental selection of forecasting models. Journal of Operations Management.

Prak, D., Teunter, R., 2018. A general method for addressing forecasting uncertainty in inventory models. International Journal of Forecasting. In press.

Prak, D., Teunter, R., Syntetos, A., 2017. On the calculation of safety stocks when demand is forecasted. European Journal of Operational Research 256 (2), 454 – 461.

Sagaert, Y. R., Kourentzes, N., De Vuyst, S., Aghezzaf, E.-H., Desmet, B., 2018. Incorporating macroeconomic leading indicators in tactical capacity planning. International Journal of Production Economics.

Sanders, N. R., Graman, G. A., 2009. Quantifying costs of forecast errors: A case study of the warehouse environment. Omega 37 (1), 116–125.

Syntetos, A. A., Babai, M. Z., Gardner Jr, E. S., 2015a. Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. Journal of Business Research 68 (8), 1746–1752.

Syntetos, A. A., Babai, M. Z., Jr., E. S. G., 2015b. Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. Journal of Business Research 68 (8), 1746 – 1752, special Issue on Simple Versus Complex Forecasting.

Syntetos, A. A., Boylan, J. E., 2005. The accuracy of intermittent demand estimates. International Journal of forecasting 21 (2), 303–314.

Syntetos, A. A., Boylan, J. E., 2006. On the stock control performance of intermittent demand estimators. International Journal of Production Economics 103 (1), 36–47.

Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., 2010. Judging the judges through accuracy-implication metrics: The case of inventory forecasting. International Journal of Forecasting 26 (1), 134–143.

Teunter, R. H., Duncan, L., 2009. Forecasting intermittent demand: a comparative study. Journal of the Operational Research Society 60 (3), 321–329.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.

Xia, Y., Tong, H., et al., 2011. Feature matching in time series modeling. Statistical Science 26 (1), 21–46.