

# Incorporating macroeconomic leading indicators in tactical capacity planning

Yves R. Sagaert<sup>a,c,\*</sup>, Nikolaos Kourentzes<sup>b</sup>, Stijn De Vuyst<sup>a</sup>, El-Houssaine Aghezzaf<sup>a</sup>, Bram Desmet<sup>c</sup>

<sup>a</sup>*Department of Industrial Management, Ghent University, Technologiepark 903, Zwijnaarde 9052, Belgium*

<sup>b</sup>*Department of Management Science, Lancaster University Management School, Lancaster, LA1 4YX, UK*

<sup>c</sup>*Solventure NV, Sluisweg 1, Gent 9000, Belgium*

---

## Abstract

Tactical capacity planning relies on future estimates of demand for the mid- to long-term. On these forecast horizons there is increased uncertainty that the analysts face. To this purpose, we incorporate macroeconomic variables into microeconomic demand forecasting. Forecast accuracy metrics, which are typically used to assess improvements in predictions, are proxies of the real decision associated costs. However, measuring the direct impact on decisions is preferable. In this paper, we examine the capacity planning decision at plant level of a manufacturer. Through an inventory simulation setup, we evaluate the gains of incorporating external macroeconomic information in the forecasts, directly, in terms of achieving target service levels and inventory performance. Furthermore, we provide an approach to indi-

---

\*Yves R. Sagaert

*Email address:* `yves.sagaert@ugent.be` (Yves R. Sagaert)

cate capacity alerts, which can serve as input for global capacity pooling decisions. Our work has two main contributions. First, we demonstrate the added value of leading indicator information in forecasting models, when evaluated directly on capacity planning. Second, we provide additional evidence that traditional metrics of forecast accuracy exhibit weak connection with the real decision costs, in particular for capacity planning. We propose a more realistic assessment of the forecast quality by evaluating both the first and second moment of the forecast distribution. We discuss implications for practice, in particular given the typical over-reliance on forecast accuracy metrics for choosing the appropriate forecasting model.

*Keywords:* Leading indicators, Inventory, Forecasting, Capacity planning

---

## 1. Introduction

In a manufacturing context, production needs to be planned well in advance, so that production capacity can be pooled on a global scale. A key requirement in tactical capacity planning is to anticipate the future total demand and ensure availability of adequate production capacity. Volling et al. (2013) emphasises that the major challenge in this is to identify the right level of tactical flexibility within a company. This flexibility translates into a trade-off between investing in additional capacity and the deployment of capacity adjustment options. A necessary input for the tactical decision making process are reliable demand predictions for the future periods, which must also accommodate for the associated supply chain risks (Stadtler, 2005). The

latter are typically handled by calculating safety stock requirements, which in turn are based on the variability of the forecast errors (Nahmias and Cheng, 2009), as the future demand is unknown and must be predicted.

Therefore, the total capacity requirements are conditional on demand forecasts, accounting for forecast error variability. Yet, the selection of forecasting models is typically done on accuracy metrics (Fildes, 1992, Ord et al., 2017), which is a myopic choice as they consider solely the first moment of the error distribution and ignore higher moments, which can have significant implications for decision making, for instance in inventory management (Barrow and Kourentzes, 2016). In this paper, we address this issue by performing an inventory simulation to estimate the demand risk, accommodating specific company practices, such as the inventory policy. We contrast the performance of forecasting models selected in the conventional way, by assuming that forecast accuracy is an adequate proxy for evaluating the quality of predictions for capacity planning decisions, and by the proposed inventory simulation based evaluation, which estimates the demand risk explicitly.

Using inventory performance metrics is in line with the literature that argues forecasting models used for inventory management should be evaluated as such, particularly for slow moving items (Babai et al., 2013, Kourentzes, 2013, Wang and Petropoulos, 2016). The motivating difference in this work is the problem context. In contrast to operational inventory decisions that are short-term, tactical capacity planning requires longer term forecasts that have to account for potential changes in the market dynamics. The majority

of supply chain management forecasting is based on extrapolative methods, which model demand patterns from the past, with the potential superposition of market information by experts in the form of judgemental adjustments (Fildes et al., 2009). This faces two problems, experts have inconsistent performance (Trapero et al., 2013) and extrapolative forecasts are incapable of capturing changes in market dynamics, as they are based only on past historical observations of the target variable. Furthermore, it could be argued that using experts can be an expensive and labour intensive task that may not always be desirable or even feasible. Recent advances in tactical forecasting propose using leading indicator information, such as macroeconomic variables that can provide early indications of changes in market dynamics. Sagaert et al. (2018) argued that including external macroeconomic leading indicators can be beneficial in tracking changes in the business environment, leading to improved tactical-strategic forecasts.

This paper extends this research by investigating the usefulness of leading macroeconomic indicators for tactical capacity planning, by including this information on the plant level. Notably, the literature on assessing forecast quality for capacity planning is sparse and our findings highlight both successful forecasting models for this purpose and provide a valid evaluation framework. We perform our analysis on a real case of a manufacturer with multiple plants in different countries and contrast the use of state-of-the-art extrapolative forecasting methods and models using leading indicators. We demonstrate that should one choose the best forecasts based on accuracy

metrics, a suboptimal solution could be reached, as the supply chain risk would not be accounted for fully. On the other hand, using the proposed evaluation demonstrates that leading indicators enhance forecasts in a variety of circumstances. Therefore, the contribution of the paper is two-fold: (i) we demonstrate the usefulness of leading indicators in forecasts for tactical capacity planning, and propose a model that identifies and uses them effectively and automatically; and (ii) we demonstrate the benefits of considering supply chain risk in selecting the appropriate forecasting model.

The rest of the paper is organised as follows. Section 2 provides a brief literature review, motivating the research questions. Section 3 provides the theoretical formulations on experimental setup. Section 4 presents the empirical data and the results, and Section 5 contains concluding remarks.

## **2. Background**

The majority of capacity planning work has looked at the operational side of the problem (for example, Huang et al., 2008, Lee and Kim, 2002). Studies focus on optimising the production schedule and inventory rules, assuming the forecasts as given (Goyal and Giri, 2003), or overly simplistic (Silver et al., 2016). A literature review by Volling et al. (2013) notes that most models for the planning of tactical capacity assume deterministic demand and focus on a single production site. He identifies a gap of stochastic approaches in the context of global planning, and expects that improvements can be obtained by models that incorporate stochastic influences to assess the risk that is

inherent to mid- to long-term tactical planning decisions. These risks can originate from internal effects within the company (e.g. product demands, product life cycles, market prices or production costs and transport costs; Vidal and Goetschalckx, 2000) or external effects (e.g. exchange rates Meyer, 2004, political decisions or other macroeconomic factors).

For the long-term strategic capacity planning, Chien et al. (2017) addresses demand uncertainty via judgemental forecasting for a semiconductor case. Kourentzes et al. (2014) and Athanasopoulos et al. (2017) demonstrate that conventional extrapolative forecasting methods perform poorly for such long-term predictions as both the selection and parametrisation of the models are based on short-term focused statistics. Bihlmaier et al. (2008) includes discrete probability distributions of the future demand into their strategic network planning for an automobile industry case. However, this research only allows for flexible capacity by closing existing or open new production sites, as these are the main decisions that are taken on a strategic planning level (Goetschalckx, 2002).

However, a tactical decisions process focusses on where to manufacture which product or product family and from where to satisfy the customers demand. In this paper, we focus on this tactical decision problem, where having a reliable forecast is more challenging than in an operational setting due to the horizons involved, the forecast is the primary input for effective decision making.

The effects of demand uncertainty and forecast errors on supply chain

planning models have received some attention in the literature. Fildes and Kingsman (2011) demonstrate that not accounting for the stochastic risk results in very poor supply chain performance. They also found that in manufacturing problems that exhibit high demand uncertainty and consequently forecast error, any improvements in accuracy should lead to substantial percentage improvements in unit costs, by eliminating waste and inefficient planning. Enns (2002) looked at the effect of forecasting bias and demand uncertainty on production planning and found mixed effects. He argued that if bias is unavoidable, it is preferable to under-forecast and handle the demand uncertainty and associated supply chain risk by using safety stock, instead of being biased towards over-forecasting.

Estimating the uncertainty around a point forecast is indispensable for tactical capacity planning and inventory control. This can be assessed by reviewing the distribution of forecast errors. Yet, to the best of our knowledge the measures of forecast uncertainty have been neglected in capacity planning research. Barrow and Kourentzes (2016) looked at the impact of the distribution of forecast errors on safety stock. They found that more accurate forecasts, as constructed by combining different forecasts, resulted in more symmetric distributions that demonstrated more consistent in- and out-of-sample errors. Naturally, this led to lower safety stock requirements as the risk was captured better. Kolassa (2016) focussed on evaluating forecasts of count data in retail, and argued in favour of considering the complete distribution instead of summary accuracy metrics. Wang and Petropoulos (2016)

compared different statistical, managerially adjusted forecasts and combinations of them in terms of forecast accuracy, bias and inventory performance, concluding that inventory variance is positively affected by the variance of the forecast errors. Babai et al. (2013) contrasted the benefit of information sharing on both forecast accuracy and inventory performance.

It is well established that enhancing extrapolative forecasts with additional relevant information can result in better behaved and narrower forecast error distributions. The type of relevant information depends on the forecasting objective, for example, in an operational setting including promotional (Kourentzes and Petropoulos, 2015, Ma et al., 2016) or point-of-sales (Trappero et al., 2012) information can substantially improve forecasts. At a tactical level, because of the geographical constraints on manufacturing capacity and inventory, the supply chain is strongly affected by external changes (Prater et al., 2001). Research by Chien et al. (2017) states that leading indicators with information from an economic environment aid domain experts in formulating long-term demand forecasts.

This motivates us to quantify the impact on sales forecasting with external macroeconomic indicators on the supply chain. This is supported by previous research that has argued the benefits of using such external information to improve strategic forecast accuracy (Sagaert et al., 2017). While macroeconomic indicators have been used successfully in forecasting macroeconomic variables (Bai and Ng, 2008, Stock and Watson, 2012, Bulligan et al., 2015), in this paper, we want to assess their value on tactical capacity planning

level.

There is a clear gap in the literature in exploring appropriate forecasting models for tactical capacity planning and their impact on the required capacity decisions, or on the production plan and the inventory on mid-term.

### **3. Methodology**

#### *3.1. Tactical capacity planning*

Capacity planning on a tactical level is concerned with the best allocation of production plant capacity to a production plan on a global scale. Volling et al. (2013) distinguishes two main elements in tactical capacity: technological capacity and organisational capacity. The capacity problem includes then human resources constraints, actual machine availability, as well as internal company policies on a global level. At an operational level, capacity planning can be myopic, due to the short-term focus, and does not guarantee available resources over longer periods of time. Tactical capacity planning enables to overcome this problem, by taking the capacity constraints into account on mid-term planning. The major advantage of tactical capacity planning is that, the capacity of different production plants can still be reallocated on a medium to long horizon without causing a large impact on the operational planning. Naturally, to achieve this accurate tactical forecast are required.

#### *3.2. Indicator forecasting model*

Sagaert et al. (2018) proposes a methodology for incorporating leading indicators to sales forecasting on a strategic-tactical level, based on the Least

Absolute Shrinkage and Selection Operator (LASSO) methodology (Tibshirani, 1996). Macroeconomic leading indicators are automatically identified, together with their lead order, i.e. how many periods in advance the indicator provides leading information to explain the movement of sales. The model can simultaneously select information from the univariate time series information (e.g. seasonality) as well as external information.

When using exogenous variables in a forecasting context, their future values will typically be unknown. To overcome this, we formulate unconditional forecasts, that is we only use information available up until the forecast origin at period  $t$ . This can be achieved by lagging the exogenous variables accordingly, so that only observed values are used as inputs for the forecasts. Given  $h$  the forecast horizon, and  $k$  the lag order for an indicator, the leading effect should be long enough so that  $k \geq h$ . This ensures that we never require information that is not observed. Naturally, for each forecast horizon a different constraint on  $k$  is imposed and therefore we construct  $h$  different forecasting models to account for this.

Let  $\hat{Y}_{t,h}$  be the forecast of the demand in period  $t+h$ , made at  $t$ . We can incorporate leading indicator information to the forecast using a regression formulation:

$$\hat{Y}_{t,h} = \beta_0 + \sum_{s=1}^{S-1} \beta_s D_{t,s} + \sum_{j=1}^P \sum_{k=h}^L \beta_{jk} X_{j,t-k}, \quad (1)$$

where  $\beta_0$  is a constant,  $\beta_s$  is the coefficient for the binary seasonal dummy  $D_{t,s}$  for seasonal period  $s$ .  $\beta_{jk}$  is the coefficient for the  $j = 1, \dots, P$  indicator

$X_{j,t-k}$ , is lagged by  $k$  periods in time. Observed that (1) has two groups of terms: the first captures seasonality, and the latter consists of the leading indicators.

The model parameters are optimised using the LASSO cost function:

$$\sum_{i=1}^N \left( Y_i - \sum_{s=1}^{S-1} \beta_s D_{i,s} - \sum_{j=1}^P \sum_{k=h}^L \beta_{jk} X_{j,i-k} \right)^2 + \lambda \left( \sum_{s=1}^{S-1} |\beta_s| + \sum_{j=1}^P \sum_{k=h}^L |\beta_{jk}| \right), \quad (2)$$

where  $Y_i$ ,  $i = 1, \dots, N$ , are the in-sample observations. These, together with the seasonal dummies ( $D_{i,s}$ ) and the lagged indicators ( $X_{j,i-k}$ ) are normalised to zero mean and unit standard deviation. This ensures that the scale of the various indicators does not bias the estimation. The scalar  $\lambda$  controls the amount of shrinkage. In (2) the first term measures the fit using the standard regression quadratic cost, while the second term penalises the fit by the absolute size of the various coefficients. This forces them to become small, and due to the absolute loss even excluded from the model. Therefore, the LASSO estimates the predictor coefficients and makes a selection of the most relevant ones. The model is able to select among the predictors, even if  $PL \gg N$ . When  $\lambda$  is set to zero, all predictors enter the model and becomes equivalent to conventional ordinary least squares regression. The appropriate  $\lambda$  is selected using cross-validation. Finally, once all coefficients have been

estimated the constant  $\beta_0$  is calculated as:

$$\beta_0 = \frac{1}{N} \sum_{i=1}^N Y_i - \left( \sum_{s=1}^{S-1} \beta_s \frac{1}{N} \sum_{i=1}^N D_{i,s} + \sum_{j=1}^P \sum_{k=h}^L \beta_{jk} \frac{1}{N} \sum_{i=1}^N X_{j,i-k} \right). \quad (3)$$

The potential pool of macroeconomic indicators is immense. Sagaert et al. (2018) showed that LASSO is capable to effectively select useful and relevant indicators automatically. Note that the selection problem is substantial, as one has to consider all relevant indicators in all their lagged realisations as potential inputs. They also showed that the model can take advantage of expert information, in which case experts can do a coarse pre-filtering for relevant types of variables, resulting in improved forecasting performance.

### 3.3. Estimation of forecast uncertainty

The demand process follows an unknown stochastic model, associated with some uncertainty. Based on the observed values  $Y_1, \dots, Y_t$ , at period  $t$ , the model then proposes a random variable  $\hat{Y}_{t,h}$  as the projected demand  $h$  periods later, with distribution  $F_{t,h}(x) = \text{Prob}[\hat{Y}_{t,h} \leq x]$ .

In the literature two approaches have been advocated to assess the uncertainty of the stochastic process. One is suggesting to use the variance of demand to estimate this uncertainty (for example, see Heizer and Render, 2004), while the other suggests the variability of the forecast error (for example, Silver et al., 2016). Although there are cases that these two may converge, the latter is more general. This is easy to exemplify considering a deterministic seasonal demand pattern, the variance of which is non-zero, yet

it implies no uncertainty and an appropriate forecasting model would result in zero errors. Therefore here we follow the second approach.

Nonetheless, in estimating the uncertainty for multi-step-ahead forecasts there is no single approach. Given a forecasting model, the expectation of  $\hat{\mathbf{Y}}_{t,h}$  is the point forecast  $\hat{Y}_{t,h}$ , which is associated with a forecast error  $e_{t,h}$ . We typically assume that the error is Gaussian with zero mean and standard deviation  $\sigma$ . When a forecasting model is available, typically there are analytical expressions of  $\sigma$  for multi-step forecasts (for example Hyndman et al., 2008, provides such for the exponential smoothing family of models), which require an estimate of the observed one-step ahead forecast error  $\hat{\sigma}_1$ . This is typically calculated as:

$$\hat{\sigma}_1 = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_{i-1,1})^2}, \quad (4)$$

on the available  $N$  historical demand observations and is the well known Root Mean Squared Error (RMSE). Although there is research that has demonstrated that estimating  $\sigma_1$  on the in-sample data will underestimate the size of the uncertainty, as the same sample is used to parametrise the forecasting model (Barrow and Kourentzes, 2016), it is commonly used to derive estimates of the multi-step error variance.

However, in the case of the model (1) that utilises the leading indicator information due to the unconditional forecasting setup, a different model is produced for each forecast horizon  $h$ . Therefore, we can estimate directly

the standard deviation for each respective horizon,  $\hat{\sigma}_h$  directly from each of the  $h$  models as follows:

$$\hat{\sigma}_h = \sqrt{\frac{1}{N-h+1} \sum_{i=h}^N (Y_i - \hat{Y}_{i-h,h})^2}. \quad (5)$$

Note that the theoretically calculated  $\sigma_h$  typically inflates for increasing  $h$ , depending on the model form. This is imposed by the iterative nature of how forecasts are produced, i.e. the forecast for period  $t+h-1$  is used to produce the  $t+h$  period forecast, any errors that occur in periods preceding  $t+h$  propagate to that period, introducing covariances, thus inflating the expected uncertainty, which leads to the commonly assumed behaviour that forecast uncertainty increases for multi-step forecasts. This is not the case for the leading indicator model. For each horizon a different model is used and no propagation of errors occurs. Therefore, a further advantage of the proposed forecasting approach is that it avoids the aforementioned introduction of covariances, retaining the forecast uncertainty to lower levels.

#### 3.4. Inventory simulation

We design an inventory simulation, which allows us to simulate the production over a rolling origin and evaluate the evolution of the inventory, comparing the effect of different forecasts as input. Our assumptions for the inventory process are the following:

- The inventory of a particular (group of) item(s) has a periodic review

at fixed periods of 1 month.

- Production of items is Make-to-Stock and happens continuously throughout the period, but observing the inventory position and satisfying demand with items from the inventory is done once at the end of every period.
- In our context, the inventory is controlled by the decision of how much of those items to produce in each period. A variant of the classical service level method, see e.g. Thomopoulos (2015), is used which involves solving a so-called newsboy problem in every period. As a complication however, there is a production planning stand-off of  $g - 1$  periods which means that at period  $t$  (i.e. at the end of period  $t$ ) the quantities to be produced in periods  $t + 1$  to  $t + g - 1$  are already committed in previous periods and can no longer be altered.

Figure 1 visualises the production environment in monthly buckets.

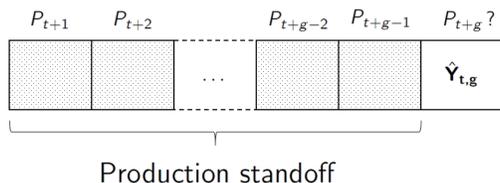


Figure 1: Allocation of the production in period  $t + g$ , to be decided at the end of period  $t$ , in case of a production standoff of  $g - 1$  periods.

Let  $P_t$  be the amount of items produced in period  $t$  and the inventory position (on hand minus backorders) at the end of period  $t$  is observed to be  $I_t$ . As in

previous sections, let  $Y_t$  be the actual demand in period  $t$  and  $\hat{Y}_{t,h}$  the random variable in period  $t+h$  as forecasted at period  $t$ , with the point forecast  $\hat{Y}_{t,h}$  and standard deviation  $\hat{\sigma}_{t,h}$ . At  $t$ , the future production quantities  $P_{t+1}$  to  $P_{t+g-1}$  are already established so that the inventory position at  $t+g-1$  can be estimated as

$$\hat{I}_{t,g-1} = I_t + \sum_{h=1}^{g-1} (P_{t+h} - \hat{Y}_{t,h}). \quad (6)$$

The demand at  $t+g$  will be satisfied entirely whenever

$$I_{t+g-1} + P_{t+g} > Y_{t+g}. \quad (7)$$

If we choose to produce  $P_{t+g}$  items, the probability that the demand in period  $t+g$  is not entirely satisfied can be estimated as

$$\begin{aligned} \hat{\alpha}(P_{t+g}) &= \text{Prob}[\hat{I}_{t,g-1} + P_{t+g} < \hat{Y}_{t,g}] \\ &= \text{Prob}[I_t + \sum_{h=1}^{g-1} (P_{t+h} - \hat{Y}_{t,h}) + P_{t+g} < \hat{Y}_{t,g}] \\ &= \text{Prob}[\sum_{h=1}^g \hat{Y}_{t,h} > P_{t+g} + I_t + \sum_{h=1}^{g-1} P_{t+h}]. \end{aligned} \quad (8)$$

Given the (continuous) distribution function

$$F_{t,g}^*(x) = \text{Prob}[\sum_{h=1}^g \hat{Y}_{t,h} \leq x], \quad (9)$$

of the summed forecasted demand and a certain acceptable shortage proba-

bility  $\alpha_{t+g}$  for period  $t + g$ , the production quantity  $P_{t+g}$  in that period can be allocated as

$$P_{t+g} = F_{t,g}^{*-1}(1 - \alpha_{t+g}) - I_t - \sum_{h=1}^{g-1} P_{t+h}, \quad (10)$$

The challenge therefore is to obtain the distribution function (9) from the forecasting model. Currently the forecasting models do not account for correlation in the demand sequence  $Y_1, \dots, Y_t$ . In fact, it is assumed that the forecasts  $\hat{Y}_{t,h} \sim N(\hat{Y}_{t,h}, \hat{\sigma}_{t,h}^2)$ ,  $h > 0$ , form an independent sequence so the distribution function  $F_{t,g}^*(x)$  is again that of a normal random variable

$$\sum_{h=1}^g \hat{Y}_{t,h} \sim N\left(\sum_{h=1}^g \hat{Y}_{t,h}, \sum_{h=1}^g \hat{\sigma}_{t,h}^2\right). \quad (11)$$

The above proposes a sequential (period-by-period) way of allocating the production, assuming there are no constraints on how much the company can produce in a period. Unless items in stock can be removed by means other than customer demand (e.g. transfer to other plants), an obvious constraint on the production would be that it is nonnegative. That is, allocation (10) becomes

$$P_{t+g} = \max\left(0, F_{t,g}^{*-1}(1 - \alpha_{t+g}) - I_t - \sum_{h=1}^{g-1} P_{t+h}\right). \quad (12)$$

This strategy aims at producing as late as possible, i.e. in the period where the demand is. However, if there would be an upper bound on the production capacity in this period  $t$  as  $P_t \leq C^{Max}$ , the current sequential ‘myopic’

planning may not be optimal or even acceptable. For example, suppose the forecast tells us now that with high certainty there will be a large demand 5 periods later which largely exceeds the upper bound on the capacity. The myopic strategy would fail to build up the inventory and prepare for this event, resulting in an unacceptably low fill rate in that period. The problem would then need to be reformulated using a finite planning horizon of  $H > g - 1$  periods (see e.g. Özer and Wei, 2004). However, as production plants in a global network do not face a strict capacity constraint, but merely a capacity pooling problem, this simulation would need to imply centralised optimisation. Yet, the decision in our case study is managerial, and plant specific. So this approach would not anticipate on the decision process in practice.

For a particular plant and forecasting model, we can simulate the resulting inventory process  $I_t$  during the forecasting test period, i.e. from period  $N + 1$  (right after the training period) to  $N + T$ . Throughout the  $T$  periods of the simulation, the actual inventory at the end of each period  $t$  is determined as

$$I_N = 0, \quad I_t = I_{t-1} + P_t - Y_t, \quad N < t \leq N + T,$$

where the production quantities  $P_t$  are planned period by period, always using the most recent available forecasts of the future demand. To judge how well a forecasting model performs, we use average on hand inventory and the fill rate (FR) achieved over the  $T$  periods as performance metrics. Recall

that fill rate refers to the average fraction of demanded items in a period which can be obtained immediately from stock, without backordering. Let  $Y_t^{(p)}$  and  $I_t^{(p)}$  respectively be the actual demand in period  $t$  and the inventory position at the end of period  $t$  for plant  $p$ . Then

$$\bar{I}^{+(p)} = \frac{1}{T} \sum_{t=n+g}^{n+T} \max(0, I_t^{(p)}), \quad (13)$$

and

$$\text{FR}^{(p)} = \frac{\sum_{t=n+g}^{n+T} \max(0, Y_t^{(p)} + \min(0, I_t^{(p)}))}{\sum_{t=n+g}^{n+T} Y_t^{(p)}}. \quad (14)$$

are the average on hand inventory and the fill rate (FR) estimated during the test period. Note that we left out the first  $g - 1$  periods in these estimates.

## 4. Empirical Evaluation

### 4.1. Case Study

The dataset contains real sales data from 5 global plants of a tire manufacturer for the period of 2005-2015 on a monthly frequency. This data is not adjusted for any seasonal or calendar effects. Two of these plants only started in April, 2010 and data starts from this point onwards. The period on which the models are fit is 2005:01-2012:12 and the test period is 2013:01-2014:12. The time series are found to be trended and seasonal, as verified by statistical tests and visual inspection. In our experiments, we forecast 6 months ahead, in a rolling origin setting. In the company production, there is

a production stand-off of 5 months, so forecasts of  $h = 6$  months are needed to plan production.

The set of potential leading indicators consists of 1,011 monthly variables and is the result of a coarse pre-filtering by one of the plant managers. All external data is retrieved from the Federal Reserve Economic Data (FRED). The indicators consist of different types of data, covering different aspects of the macroeconomic dynamics. Table 1 provides examples of some of these groups.

The optimal leading effect of the indicators is determined by LASSO. For this, the indicators need to be lagged in time, at every potential relevant lag (1 to  $L$ ) prior to inputting them to the LASSO. This enlarges the set of indicators by a factor  $L$ . In this paper, we accommodate for leading effects up to one year ahead and set  $L = 12$ . In order to forecast  $h = 1$ , the model needs to select the most relevant predictor from a set of 12,132 shifted indicators and 11 seasonal dummies. For  $h = 2$ , the set becomes smaller, amounting to 11,121 lagged indicators and 11 seasonal dummies, which is necessary due to the unconditional setup of the LASSO model, as outlined in section 3.2.

#### *4.2. Benchmark models*

We compare the forecast model using exogenous indicators with three traditional forecasting models. First, the Naive model (or Random Walk) assumes that the sales history is uninformative, and constructs a forecast that is equal to the last observation for the next  $h$  periods. Note that since the

Table 1: Examples of indicators clusters and their typical units.

Cluster of indicators	Unit
Total inventories value at Manufacturers	Million of Dollars
Total shipments value at Manufacturers	Million of Dollars
Passenger car registrations	Index
Diffusion index of national activity	Index
National index of consumer prices for fuel	Index
National index of consumer prices for electricity	Index
National index of consumer prices for passenger transport	Index
Consumer price index for motor vehicle parts	Index or Growth Rate Previous Period
Domestic auto production volume	Thousands of Units
Motor vehicle retail sales	Thousands of Units
Business and consumer motor vehicle loans outstanding	Millions of Dollars
Import and export of fuels	National currency or Index
Unemployment level	Thousands of Persons
Personal income	Billions of Dollars

Naive requires no parameters to be estimated, or input from the modeller, it is a fundamental benchmark that more complex forecasts should out-perform to be of value.

Second, we use the case company current forecast as a benchmark, which is the Holt-Winters method. For this evaluation, we implement Holt-Winters within the state-space exponential smoothing family of models (Hyndman et al., 2002) that provides both likelihood expression to optimise the model parameters and analytical expressions for multi-step error variance.

Third, we allow for the complete exponential smoothing (ETS) family of models, choosing the most appropriate one by Akaike Information Criterion corrected for sample size (AICc). Given the widespread use of exponential

smoothing models in supply chain forecasting, we do not present the different models here and refer the reader to Hyndman et al. (2008) for the details. Nonetheless, we note that the complete family of models allows for stochastic trend and seasonality of various forms, and the error term of the model may interact additively or multiplicatively, resulting in a very wide number of time series that it performs well. In fact, Gardner (2006) reports the good performance and widespread use of the models in both academia and practice.

The motivation behind our benchmark choices is to assess the added value of incorporating leading indicators over well established univariate forecasting models, as well as the current company forecast, to demonstrate any added value of the more complex model.

#### 4.3. Performance measures

We evaluate the performance of the competing models on forecast accuracy, estimated uncertainty and on inventory performance. From an accuracy standpoint, we use the Average Relative Root Mean Squared Error (AvgRelRMSE):

$$\text{AvgRelRMSE}_h = \sqrt[5]{\prod_{p=1}^5 \left( \frac{\text{RMSE}_h^A}{\text{RMSE}_h^B} \right)}, \quad (15)$$

and

$$\text{RMSE}_h = \sqrt{\frac{1}{N} \sum_{t=1}^N \left( Y_{t+h} - \hat{Y}_{t,h} \right)^2}. \quad (16)$$

The  $\text{AvgRelRMSE}_h$  is the geometric mean across all five plants (p) of the RMSE ratio of a forecast over a benchmark, across all evaluation sample,

for a given horizon  $h$ . We use ETS as the benchmark forecast, given its good performance in the literature. If a forecast outperforms ETS, then  $\text{AvgRelRMSE}_h$  becomes smaller than one and vice-versa. Furthermore, we can calculate percentage improvements in accuracy over the benchmark as  $1 - \text{AvgRelRMSE}_h$ .

This metric is based on the Average Relative Mean Absolute Error (AvgRelMAE), proposed by Davydenko and Fildes (2013). The AvgRelMAE has been shown to have desirable statistical properties: it does not exhibit the bias common in percentage based errors; it is scale independent, allowing us to summarise across the different plants; it is robust to calculation issues; and it is easy to interpret. We use RMSE as a basis so as to retain a direct connection with forecast uncertainty and inventory.

We use a similar metric to compare the estimated forecast uncertainty between forecasts, as captured by  $\hat{\sigma}_h$  (section 3.3). We define  $\text{AvgRel}\hat{\sigma}_h$  as:

$$\text{AvgRel}\hat{\sigma}_h = \sqrt[5]{\prod_{p=1}^5 \left( \frac{\hat{\sigma}_h^A}{\hat{\sigma}_h^B} \right)}, \quad (17)$$

where  $\hat{\sigma}_h^A$  refers to the evaluated forecast and  $\hat{\sigma}_h^B$  to the benchmark that is again ETS. The values of  $\text{AvgRel}\hat{\sigma}_h$  are read in the same way as AvgRelRMSE, but their key difference is that the former is based on in-sample estimates of uncertainty, while AvgRelRMSE is calculated on the test set.

Finally to assess inventory performance, a weighted average is taken

across the different plants:

$$\bar{I}_\alpha^+ = \sum_{p=1}^5 w_p \bar{I}^{+(p)}, \quad \text{and} \quad \overline{FR}_\alpha = \sum_{p=1}^5 w_p FR^{(p)}, \quad (18)$$

with weights according to the actual demanded volumes per plant

$$w_p = \frac{\sum_{t=n+1}^{n+T} Y_t^{(p)}}{\sum_{p=1}^5 \sum_{t=n+1}^{n+T} Y_t^{(p)}}. \quad (19)$$

In (18), the index  $\alpha$  refers to the constant parameter value  $\alpha = \alpha_{t+g}$  used in the planning rule (12).

#### 4.4. Results

##### 4.4.1. Forecasting accuracy

The accuracy results for AvgRelRMSE for each forecast horizon are presented in table 2. The results summarises the accuracy for the five plants and the most accurate result is highlighted in boldface.

Table 2: Summary AvgRelRMSE at plant level.

Forecast	Horizon						Overall
	1	2	3	4	5	6	
Naive	1.37	1.51	1.46	1.33	1.33	1.44	1.41
Holt-Winters	1.19	1.36	1.29	1.44	1.45	1.56	1.38
ETS	<b>1.00</b>						
LASSO	1.09	1.21	1.20	1.14	1.10	1.17	1.15

We find that ETS is consistently performing best, while LASSO that uses leading indicators is second best. The company's current forecasting model,

Holt-Winters, is substantially worse than both, particularly for longer horizons. Nonetheless, all forecasts, on average, outperform the Naive indicating that they capture some of the structure in the time series. Overall, Holt-Winters performs worse than ETS by 38% in terms of accuracy, while LASSO is 15% worse compared to ETS. Similar results were obtained by other accuracy metrics, such as AvgRelMAE and Mean Absolute Percentage Error, which are not presented here for brevity.

LASSO selects a set of leading indicators for each plant, for each horizon. These indicators originate from an indicator clusters as shown in table 1. The most frequent selected indicators are: the wholesalers in motor vehicle sales, the Consumer Price Index of the OECD group for fuel, electricity and gasoline for specific countries, the industrial production volume index of fuels and the number of people employed in specialized freight trucking. These are all relevant to the sector that the manufacturer operates in.

The accuracy results indicate that although LASSO outperforms the company benchmark, it does not add value over ETS that is considering only univariate patterns and no exogenous information. Nonetheless, as argued before, this evaluation focuses on the point forecasts and ignores the associated forecast uncertainty.

#### *4.4.2. Forecast uncertainty*

Figure 2 plots the  $\text{AvgRel}\hat{\sigma}_h$  for each model and forecast horizon. The uncertainty of ETS is always equal to 1 as it is used as basis for the calculation

of  $\text{AvgRel}\hat{\sigma}_h$ . Observe that LASSO across all horizons is associated with smaller uncertainty, effectively indicating that in-sample it has explained more variance in the series than the competing models.

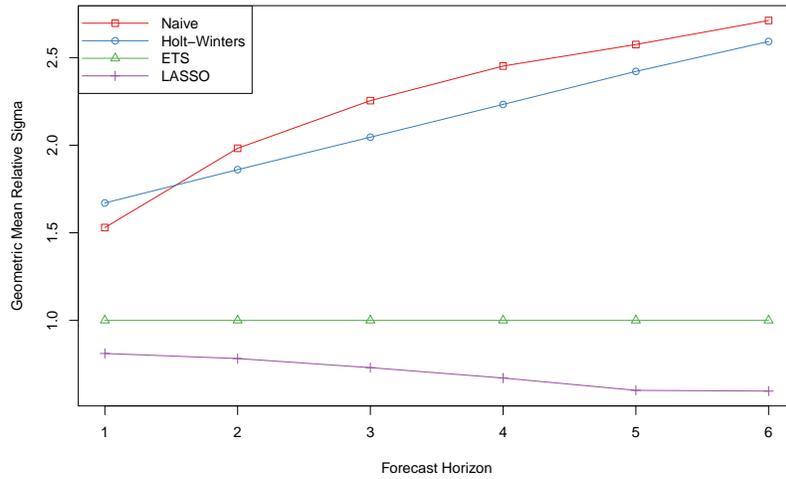


Figure 2:  $\text{AvgRel}\hat{\sigma}_h$  for the competing forecasting models.

On the other hand, the results for Naive and Holt-Winters indicate rapid deterioration over ETS as the forecast horizon increases. The uncertainty is intrinsically tied to inventory management, hence we expect these results to reflect on inventory performance.

#### 4.4.3. Service level and inventory

Figure 3 shows the model performance curves, where  $1 - \text{Fill Rate}$  is plotted against the average inventory on hand. The Naive and Holt-Winters models have substantially higher average on hand inventory. Notably the worst performer is the current company forecast, the Holt-Winters.

The inventory simulation is performed for a range of  $\alpha$  parameters, and the resulting fill rate ( $\overline{FR}_\alpha$ ) and inventory ( $\bar{I}_\alpha^+$ ) are shown in figure 3. This graph can be interpreted as follows: for a given target fill rate of the customer demand, a certain stock is required. These curves are model-specific. Under certain assumptions, a higher desired fill rate will result in a lower (1-Fill rate), which will result for all models in a higher stock on hand. Once a desired fill rate is chosen, the model then needs to be solved to estimate the parameter  $\alpha$ . This parameter  $\alpha$  represents the acceptable shortage probability, but this is a model specific parameter which needs to be optimised for each model separately. This optimisation of the  $\alpha$  happens based on the cost function, where for linearly increasing holding costs and stockout costs, we can draw the cost function visually as linear decreasing function. The overall cost minimum is then located near to the origin of the both axes. Moving along the horizontal axis will cause the stock to increase, and incur higher stock holding costs. Moving along the vertical axis will induce a lower fill rate, more shortages to the customer, and higher stockout costs. Here, we review the general cases, as the holding cost and stockout costs are company specific. We can see that the LASSO model dominates all other models in the cost minimisation.

This result contradicts the accuracy findings. The curve in figure 3 is the result of considering both the forecast uncertainty and the inventory policy of the company, on top of the point forecasts, hence closer to the relevant decision. Conversely, the accuracy results in table 2 are shown to be a weak

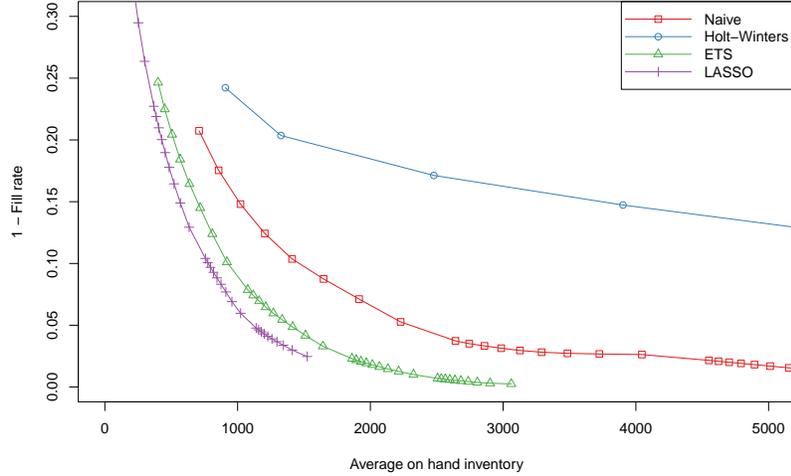


Figure 3: Performance curves of fill rate versus the average on hand inventory.

proxy in assessing the quality of the decision.

To further evaluate the quality of the forecasting models we consider the decision context. In capacity planning it is desirable to have a stable production planning. The boxplots in figure 4 represent the different production amounts  $P_t$  as determined by the inventory simulation over the test period 2013:06-2014:12. Based on this graph, we can review the production variability among the different models. The plant capacity can be visualised with a horizontal line. When the decided production amount is above that line, then the decision makers need to act in order to pool capacity on a global scale.

On the long term, the mean of the production will approximate the average demand. For all plants individually, we note that the median line in the

box plots are approximately the same over all models. However, the variability is of huge importance for capacity planning. We want the variability of the production amounts  $P_t$  to be as small as possible, as this represents a more stable production and will lead to fewer capacity problems. As figure 4 indicates, the variation on the decided production is not equal for different models. The Naive model exhibits the largest deviation in  $P_t$ . The production from the inventory simulation, based on LASSO forecasts exhibits the smallest variability. Looking at the upper quantiles of the boxplots, we can see that the capacity management will be steered better with LASSO models.

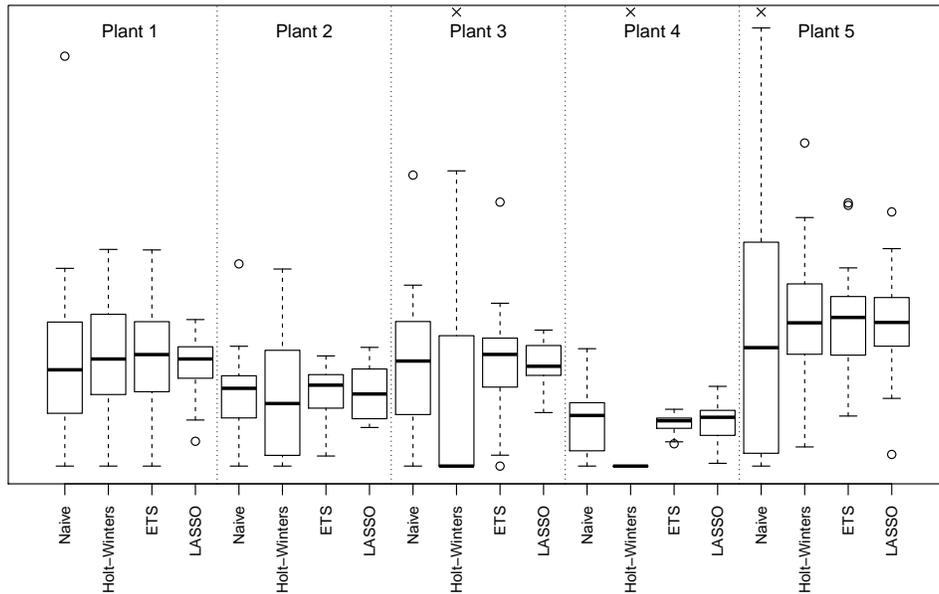


Figure 4: Box plot of different production quantities per plant. An 'o' marks an outlier that lies 1.5 IQR (Interquartile Range) away from the median. Upper 'x' denotes that there are points outside of the plot area.

## 5. Conclusion

Tactical capacity planning requires good performing forecasts. When managing capacity, production decisions are made on a mid-term level. This allows companies to perform global capacity pooling, when the available capacity on the local production plant is insufficient. The difficulty in this decision process is that mid-term forecasts typically have more uncertainty than short-term forecasts. This uncertainty can be assessed by estimating the variance of the forecast errors. This paper reviews models that make use of external information, such as macroeconomic leading indicators through direct forecasting. These models are formulated in an unconditional setup setting, which results in different forecast models for each forecast horizon. Because of this, the variances of the forecast errors have all the properties to be independent from one another. In fact, we see that these variances remain stable over time. This would indicate that formulated forecasts with external leading indicators have less uncertainty around the point forecasts, even on longer horizons.

We compared the quality of these forecasts against established univariate forecasting models, on a plant level in a company case. We demonstrate that the selection and evaluation of different forecasting models should not happen merely by evaluating the point forecasts. The entire forecast distribution should be taken into account. We evaluated the performance of different models on forecast accuracy, but also directly on measures that are important for tactical capacity planning, as service level and realised inventory. We

found that in this context forecast accuracy was a weak proxy of the forecast quality.

Therefore this paper has two contributions. First, it establishes the usefulness of leading indicators for forecasts that support mid-term capacity planning decisions. We provide a modelling framework that allows effectively selecting amongst a massive set of potential indicators, while retaining an unconditional forecast setting. Second, we present evidence that evaluating models solely on forecast accuracy is not appropriate in this context. We provide an evaluation framework to quantify the measures of importance for capacity planning decisions.

Further research should look at the effects of the reviewed models on operational levels, and the impact on SKU inventory. This research does not incorporate any judgemental available information, or benchmark against judgementally adjusted statistical forecasts, which are common in supply chain forecasting, particularly at a tactical level. Furthermore, historical stockouts truncate the distribution of demand, enlarging the systematic bias in estimates of mean and variance. Further research could extend the inventory simulation to include capacity constraints.

## References

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., and Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74.

- Babai, M., Ali, M., Boylan, J., and Syntetos, A. (2013). Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: Theory and empirical analysis. *International Journal of Production Economics*, 143(2):463–471.
- Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317. Honoring the research contributions of Charles R. Nelson.
- Barrow, D. K. and Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: Implications for inventory management. *International Journal of Production Economics*, 177:24–33.
- Bihlmaier, R., Koberstein, A., and Obst, R. (2008). Modeling and optimizing of strategic and tactical production planning in the automotive industry under uncertainty. *OR Spectrum*, 31(2):311.
- Bulligan, G., Marcellino, M., and Venditti, F. (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting*, 31(1):188–206.
- Chien, C.-F., Dou, R., and Fu, W. (2017). Strategic capacity planning for smart production: Decision modeling under demand uncertainty. *Applied Soft Computing*.
- Davydenko, A. and Fildes, R. (2013). Measuring forecasting accuracy: The

- case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3):510–522.
- Enns, S. (2002). MRP performance effects due to forecast bias and demand uncertainty. *European Journal of Operational Research*, 138(1):87–102.
- Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, 8(1):81–98.
- Fildes, R., Goodwin, P., Lawrence, M., and Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1):3–23.
- Fildes, R. and Kingsman, B. (2011). Incorporating demand uncertainty and forecast error in supply chain planning models. *Journal of the Operational Research Society*, 62(3):483–500.
- Gardner, E. S. (2006). Exponential smoothing: The state of the art - part II. *International Journal of Forecasting*, 22(4):637–666.
- Goetschalckx, M. (2002). Strategic network planning. In *Supply chain management and advanced planning*, pages 105 – 121. Springer.
- Goyal, S. and Giri, B. (2003). The production - inventory problem of a product with time varying demand, production and deterioration rates. *European Journal of Operational Research*, 147(3):549–557.

- Heizer, J. and Render, B. (2004). *Operations Management*. Number v. 1 in Operations Management. Pearson Prentice Hall.
- Huang, M.-G., Chang, P.-L., and Chou, Y.-C. (2008). Demand forecasting and smoothing capacity planning for products with high random demand volatility. *International Journal of Production Research*, 46(12):3223–3239.
- Hyndman, R., Koehler, A. B., Ord, J. K., and 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., and 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. and Petropoulos, F. (2015). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*.
- Kourentzes, N., Petropoulos, F., and Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2):291–302.

- Lee, Y. H. and Kim, S. H. (2002). Production - distribution planning in supply chain considering capacity constraints. *Computers & Industrial Engineering*, 43(12):169–190.
- Ma, S., Fildes, R., and 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.
- Meyer, B. (2004). *Value-adding logistics for a world assembly line*. Heinz Nixdorf Institut.
- Nahmias, S. and Cheng, Y. (2009). *Production and operations analysis*, volume 6. McGraw-Hill New York.
- Ord, J. K., Fildes, R., and Kourentzes, N. (2017). *Principles of Business Forecasting*. Wessex Press Publishing Co., 2nd edition.
- Özer, Ö. and Wei, W. (2004). Inventory control with limited capacity and advance demand information. *Operations Research*, 52(2):988–1000.
- Prater, E., Biehl, M., and Smith, M. A. (2001). International supply chain agility-tradeoffs between flexibility and uncertainty. *International journal of operations & production management*, 21(5/6):823–839.
- Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., and Desmet, B. (2017). Temporal big data for tactical sales forecasting in the tire industry. *Interfaces*.

- Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., and Desmet, B. (2018). Tactical sales forecasting using a very large set of macroeconomic indicators. *European Journal of Operational Research*, 264(2):558–569.
- Silver, E. A., Pyke, D. F., and Thomas, D. J. (2016). *Inventory and production management in supply chains*, volume 4. CRC Press.
- Stadtler, H. (2005). Supply chain management and advanced planning—basics, overview and challenges. *European journal of operational research*, 163(3):575–588.
- Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4):481–493.
- Thomopoulos, N. T. (2015). *Demand forecasting and inventory control*. Springer.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288.
- Trapero, J. R., Kourentzes, N., and Fildes, R. (2012). Impact of information exchange on supplier forecasting performance. *Omega*, 40(6):738–747.
- Trapero, J. R., Pedregal, D. J., Fildes, R., and Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, 29(2):234–243.

- Vidal, C. J. and Goetschalckx, M. (2000). Modeling the effect of uncertainties on global logistics systems. *Journal of Business Logistics*, 21(1):95.
- Volling, T., Matzke, A., Grunewald, M., and Spengler, T. S. (2013). Planning of capacities and orders in build-to-order automobile production: A review. *European Journal of Operational Research*, 224(2):240 – 260.
- Wang, X. and Petropoulos, F. (2016). To select or to combine? the inventory performance of model and expert forecasts. *International Journal of Production Research*, 54(17):5271–5282.