

Tactical sales forecasting using a very large set of macroeconomic indicators

Yves R. Sagaert^{a,c,*}, El-Houssaine Aghezzaf^a, Nikolaos Kourentzes^b, Bram Desmet^c

^a*Department of Industrial Management, Ghent University, Technologiepark 903, Zwijnaarde 9052, Belgium*

^b*Department of Management Science, Lancaster University Management School, Lancaster, LA1 4YX, UK*

^c*Solventure NV, Sluisweg 1, Gent 9000, Belgium*

Abstract

Tactical forecasting in supply chain management supports planning for inventory, scheduling production, and raw material purchase, amongst other functions. It typically refers to forecasts up to 12 months ahead. Traditional forecasting models take into account univariate information extrapolating from the past, but cannot anticipate macroeconomic events, such as steep increases or declines in national economic activity. In practice this is countered by using managerial expert judgement, which is well known to suffer from various biases, is expensive and not scalable. This paper evaluates multiple approaches to improve tactical sales forecasting using macro-economic leading indicators. The proposed statistical forecast selects automatically both the type of leading indicators, as well as the order of the lead for each of the selected indicators. However as the future values of the leading indicators are unknown an additional uncertainty is introduced. This uncertainty is controlled in our methodology by restricting inputs to an unconditional forecasting setup. We compare this with the conditional setup, where future indicator values are assumed to be known and assess the theoretical loss of forecast accuracy. We also evaluate purely statistical model building against judgement aided models, where potential leading indicators are pre-filtered by experts, quantifying the accuracy-cost trade-off. The proposed framework

*Yves R. Sagaert

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

improves on forecasting accuracy over established time series benchmarks, while providing useful insights about the key leading indicators. We evaluate the proposed approach on a real case study and find 18.8% accuracy gains over the current forecasting process.

Keywords: Forecasting, Tactical planning, Leading indicators, LASSO, Variable selection

1. Introduction

Sales forecasting is among the fundamental inputs for planning decisions throughout the supply chain. Estimating future demand more accurately is critical for meeting it, while minimising inventory and other related costs. These demand estimates are often modelled based on historical patterns in the data. However, including external information can improve the sales forecast performance (Currie and Rowley, 2010), especially in volatile environments. Earlier work has looked at including additional information from within the supply chain, see for example Aviv (2001), Trapero et al. (2012) and Williams et al. (2014); as well as price and promotional data (Huang et al., 2014, Ma et al., 2015). Bertrand et al. (2015) discusses the importance of weather information in retail sales. The main focus of this stream of research has been improving operational forecasts.

In contrast, tactical level dynamics can be different due to the nature of planning, the relevant horizons and the business models. Macroeconomic indicators can contain leading context information, such as changing global economic conditions. Companies review their national markets looking at the evolution and future expectations of economic indicators. These leading indicators are typically published on monthly or lower frequency, making them too slow for forecasting for operational purposes. However, for medium to long-term horizons, these macroeconomic indicators could enrich the forecasts. In several sectors tactical forecasting that supports plans for raw materials, labour, machine resources and financial planning, has a horizon of 3 to 12 months ahead. In this context macroeconomic information is relevant.

Often tactical level forecasts rely on univariate methods, which are unable to model changing conditions in a market. That forces organisations to rely on expert adjustments for this purpose, which are characterised by various biases and being unstructured (Fildes et al., 2009). In contrast to a fully statistical approach, this human interaction increases the complexity

of the forecasting process and severely limits the extent to which it can be automated.

Our objective in this paper is to propose a methodology to automatically generate forecasts for tactical horizons using relevant leading indicators and investigate the benefits for the forecasting process. We argue that leading indicators capture the driving forces of sales on a tactical level, but selecting the appropriate leading indicators and their respective lead order is not trivial. Our contribution is three-fold. First, we propose a framework that selects appropriate indicators, amongst tens of thousands, and copes with the unknown future values of the selected indicators. The framework selects the appropriate lead order of each indicator and takes into account typical constraints of tactical forecasting for supply chain management. Second, we compare fully statistical selection of indicators against human-aided selection, where human experts pre-select useful indicators from which the statistical model subsequently identifies the final set of useful indicators. Third, we assess the theoretical loss of accuracy between realistic forecasting and if we were to assume that the future values of indicators were known.

We evaluate our framework using a case study of tactical forecasting of a major supplier to a global tire manufacturer in the US and Europe. Their Sales and Operation Planning (S&OP) cycle requires a 12 month forecast to plan labour, machine capacity and raw materials, with the latter having a lead time of 4 to 6 months. We use a pool of 67,851 macro-economic indicators to enhance the current forecasting process.

We find that leading indicators can improve the final forecast both in terms of accuracy and insights, despite the relatively short history of sales and massive set of potential indicators. Our analysis also suggests that human experts can add value in pre-identifying relevant broad groups of indicators, from which a smaller set is subsequently identified using a statistical model.

The rest of the paper is organised as follows. In the next section, we review relevant literature. In section 3, we discuss the proposed methodology to perform the selection of the macroeconomic variables and incorporate them into a forecast model. In section 4, we introduce the case study, outline the design of experiment and benchmark models. Findings are presented in section 5 with concluding remarks following in section 6.

2. Exogenous indicators in the literature

The potential of including external indicators has been explored in sales forecasting, but limited work has been done on the level of tactical sales forecasting. Nonetheless, there is a clear interest from industry. The surveys of Dalrymple (1987) and Klassen and Flores (2001) report that 38% and 44% of industrial firms respectively are already using leading indicators. Both studies conclude that leading indicators are mostly used on longer strategic forecasting horizons. Sanders and Manrodt (2003) show that 56% of the companies in their sample consider leading indicators being important for forecast formulation. Although these surveys indicate that companies see value in leading indicators, none discuss how many indicators are considered, or whether any statistical modelling is used. Furthermore, organisations indicate that there is a lack of methods that are capable of leveraging external data, which is currently typically done at a strategic level by judgemental means, even if this compromises forecast accuracy (Weller and Crone, 2012).

Including leading indicators in a forecasting model introduces two modelling stages. The first stage is the selection of the appropriate leading indicators from the complete set of potential ones. The second stage models their importance and impact accordingly. Ng et al. (2008) notes that models using leading indicators introduce additional uncertainties in both stages. Significant economic indicators might be omitted, and interdependence between them can increase the modelling difficulty. Revisions of economic policies, evolution of population habits or changes of national economic structures can also alter the behaviour of macroeconomic indicators. This complicates the formulation of forecasts even further.

Intuition and expert judgement can help in selecting such variables. However, the sheer volume of potential indicators, further inflated by considering lagged realisations, makes the selection very difficult and labour intensive. This is particularly relevant when scalability and automation of forecasting is desirable. Nevertheless, incorporating market intelligence and other available information has great potential to improve the sales forecasts (Fildes et al., 2009). To assess these market insights, forecasting experts base their judgement often on the economic expectations, numerically represented in economic indicators (Lawrence et al., 2000). However, Leitner and Leopold-Wildburger (2011) show that human experts make inefficient use of additional information, such as leading indicators, making the forecasts less accurate. Petropoulos et al. (2015) adds that forecasters can repeatedly make the same

erroneous adjustments, based on inaccurate information. They persist in their belief that large changes to the statistical forecast are needed, even when forecasting accuracy evidence shows otherwise.

Trapero et al. (2013) argues that classes of managerial adjustments of univariate statistical forecasts can be systematised through statistical modelling, leading to less biased forecasts. Research on automated selection and modelling of external variables has been mostly done for operational forecasting. In several studies, external information as price and promotion improved the operational forecast accuracy of retail sales (Huang et al., 2014, Lang et al., 2015, Ali et al., 2009, Haupt et al., 2014, Trapero et al., 2014, Kourentzes and Petropoulos, 2015, Weber et al., 2017). To cope with limited historical data, Guo et al. (2013) uses prices, product information and economic indexes to improve fashion forecasts. Macroeconomic information also has been used in improving forecasting: Yap and Allen (2011) found some indicators to be relevant in forecasting turning points of Australian tourism demand, while Naser (2015) found energy consumption on a macro level contains predictive information for national economic growth. However, to the best of our knowledge, limited work has been done on including exogenous leading indicators on tactical and strategic sales forecasting that this paper aims to address.

There are several approaches to select amongst multiple potentially useful variables in the literature. Stepwise regression is very widely used and easy to implement, but it has been criticised for being likely to retain irrelevant variables (Huang et al., 2014). Another drawback of stepwise regression is that it does not scale up to big datasets. Furthermore, the stepwise selection strategy has been shown to lead to high variance solutions that can potentially overfit to the estimation sample and may have poor forecasting performance (Hastie et al., 2011). Creating individual regressions and combining their forecasts as described in Elliott et al. (2013) is computationally demanding, and gives limited insight in the driving forces of sales.

In the econometric literature, forecasting macroeconomic variables with many predictors relies mainly on Principal Component Analysis (PCA) pre-processing of the regression inputs, exemplified by Stock and Watson (1998; 2002a;b; 2006; 2012), Forni and Reichlin (1996; 1998), Bai and Ng (2002; 2006; 2008), Bai (2003), Boivin and Ng (2006), among others. By highly reducing the dimensionality of the problem using PCA, only a small number of orthogonal inputs remain as inputs. Nonetheless, most of the papers do not consider more than 500 predictors, as the initial sets of variables are

reduced judgementally. For instance, Lu et al. (2012) documents that expert opinion is used to select relevant indicators. PCA has also been used in a promotional forecasting context by Trapero et al. (2014) and Kourentzes and Petropoulos (2015). Boivin and Ng (2006) found that including more predictors to estimate the common factors makes them less useful for forecasting. In addition, PCA does not result in directly interpretable model coefficients, as these need to be post-processed to be accounted to the original variables.

An alternative approach that can cope with variable selection when the number of inputs is very large is shrinkage, often implemented using LASSO regression. LASSO has been successfully applied for operational sales forecasting with external information. Huang et al. (2014) and Ma et al. (2015) effectively dealt with the variable selection problem, measuring the impact of each variable individually. In forecasting macroeconomic variables LASSO has been found useful in selecting relevant predictors (Mol et al., 2008, Li and Chen, 2014, Bai and Ng, 2008, Bulligan et al., 2015). In recent years, the LASSO model has attracted increasingly more interest in a variety of large data problems (Tibshirani, 2011). Here, LASSO is promising because it gives direct insight in the final model, in contrast to PCA, which is valuable to the users of tactical sales forecasts. Therefore shrinkage is an attractive variable selection approach for the problem at hand.

Table 1 provides an overview of the discussed literature on several criteria: the input variable modelling approach, the context of the target variable and the type of indicator variables. The table provides the frequency of the target time series, the sample size, number of indicators, maximum order of lags considered, and the forecast horizon. Note that the listed papers use a relatively small set of external variables, and economic papers typically use long time series history. Conversely, there is a distinct tactical sales forecasting problem, where a very large set of potential predictor variables are available to choose from, with only limited training sample available from the sales history. There is very limited work in this context. This is highlighted in the classification of table 1 that indicates how different the problem characteristics of this paper are: selecting from a vast pool of indicators, with limited sample size and considering a high lead order for the input variables.

Table 1: Literature review on forecasting with external indicators

Authors	Modelling approach ^a	Target context	Input variables	Target frequency ^b	Nr. of Targets	Sample Size ^c	Nr. of Inputs ^d	Max Lag ^e	Forecast Horizons
Huang et al. (2014)	Factorising; Shrinkage	Retail	Price; Promotions	W	122	200	614	2	1; 1-4; 1-12
Lang et al. (2015)	Direct input	Retail	Price; Promotions; Store properties	W	8	89	15	-	10 (CV)
Ali et al. (2009)	Other ^f	Retail	Price; Promotions	W	168	76	100	-	25
Haupt et al. (2014)	Direct input	Retail	Price; Promotions	W	46	88	61	-	1; 4
Weber et al. (2017)	Direct input	Retail	Price; Promotions	W	648	67-85	8-9	-	17-21 (CV)
Trapano et al. (2014)	Factorising	Retail	Price; Promotions	W	148	148	26	3	-
Kourentzas and Petropoulos (2015)	Factorising; Other ^g	Retail	Promotions	W	12	104	6	-	4; 8; 12
Cuo et al. (2013)	Other ^h	Fashion retail	Price; Promotions; Product and store properties	D	3; 10	32; 35; 41; 60-75	14; 13; 9	-	3; 10
Yap and Allen (2011)	Direct Input	Tourism demand	Macroeconomic variables	Q	3	36	8	3	-
Naser (2015)	Other ⁱ	Macroeconomics	Macroeconomic variables	Y	16	46	4	4	-
Stock and Watson (1998)	Factorising	Macroeconomics	Macroeconomic variables	M	2	453	170; 224	*	12
Stock and Watson (2002a)	Factorising	Simulation;	Simulated variables;	-;	1;	100;	100-500;	*	*
Stock and Watson (2002b)	Factorising	Macroeconomics	Macroeconomic variables	M	1	480	149	*	12
Stock and Watson (2006)	Factorising	Macroeconomics	Macroeconomic variables	M	8	480	215	*	6; 12; 24
Stock and Watson (2012)	Factorising	Macroeconomics	Macroeconomic variables	M	1	540	130	*	1; 3; 6; 12
Form and Reichlin (1996)	Factorising	Macroeconomics	Macroeconomic variables	Q	143	195	109 ^j	*	1; 2; 4
Form and Reichlin (1998)	Factorising	Macroeconomics	Macroeconomic variables	Y	2	29	450 ^k	-	-
Bai and Ng (2002)	Factorising	Simulation;	Simulated variables;	-;	1;	10-8000;	10-8000;	-	-
Bai and Ng (2006)	Factorising	Macroeconomics	Macroeconomic variables	M	1	60	4883 ^m	-	-
Bai and Ng (2008)	Factorising; Shrinkage	Simulation	Simulated variables	-	4	50-100	50-200	-	4
Bai (2003)	Factorising	Macroeconomics	Macroeconomic variables	M	1	528	132	-	1; 6; 12; 24
Boivin and Ng (2006)	Factorising	Simulation	Simulated variables	-	1	50-100	25-1000	-	-
Lu et al. (2012)	Other ⁿ	Simulation;	Simulated variables;	-;	1; 2;	40-100;	20-140;	-;	-;
Ma et al. (2015)	Factorising; Shrinkage	Computer wholesaler	Calculated variables on sales	M	147	480	41; 46; 60; 147	-	12
Mol et al. (2008)	Shrinkage	Retail	Price; Promotions	W	6000	158	11	-	-
Li and Chen (2014)	Shrinkage	Macroeconomics	Macroeconomic variables	M	2	540	270-450	1	1-4
Bulligan et al. (2015)	Factorising; Shrinkage	Macroeconomics	Macroeconomic variables	M	20	600	107	4	1
Our paper	Shrinkage	Tactical supply chain demand forecasting	Macroeconomic variables	Q	5	192	60; 68; 140; 225	-	1
			Macroeconomic variables	M	4	84	67,851	12	1-12

^aWe provide 4 distinct modelling approach categories: Shrinkage of the parameters as in LASSO; Dimensionality reduction as by PCA; Direct model as in regression; Other approaches

^bTarget frequency legend: D=Daily, W=Weekly, M=Monthly, Q=Quarterly, Y=Yearly, -=Not reported

^cMultiple sample size originate from different runs

^dThe number of inputs without any lags

^eThe lags are considered of the input variables, where * stands for not specific reported and - stands for not performed.

^fRegression tree is used as modelling approach

^gMultiple Aggregation Prediction Algorithm (MAPAx) is used as modelling approach

^hHarmony search-Wrapper-based Variable Selection (HWVS) is used as modelling approach

ⁱJohansen cointegration is used as modelling approach

^jThe inputs originate from the target set and are used alternately

^kBoth target series contain 450 of cross-sectional time series, used as input variables

^lBoth target series contain 450 of cross-sectional time series, used as input variables

^mThe target series contains 4883 of cross-sectional time series, used as input variables

ⁿMultivariate Adaptive Regression Splines (MARS) is used as modelling approach

3. The proposed forecasting framework

3.1. Model inputs

We organise the information that will be included in the forecasting model in three distinct classes: (i) seasonality information; (ii) autoregressive information; and (iii) leading indicators, as shown in figure 1 that provides a flowchart of our approach. The figure shows how univariate information is combined with external information in subsequent steps. There is a central step where the option is available to insert judgemental preselection on the external data. In the last steps of the modelling framework, the time shifts 'lags' are constructed and an unconditional setup is employed, as explained in section 3.2 below.

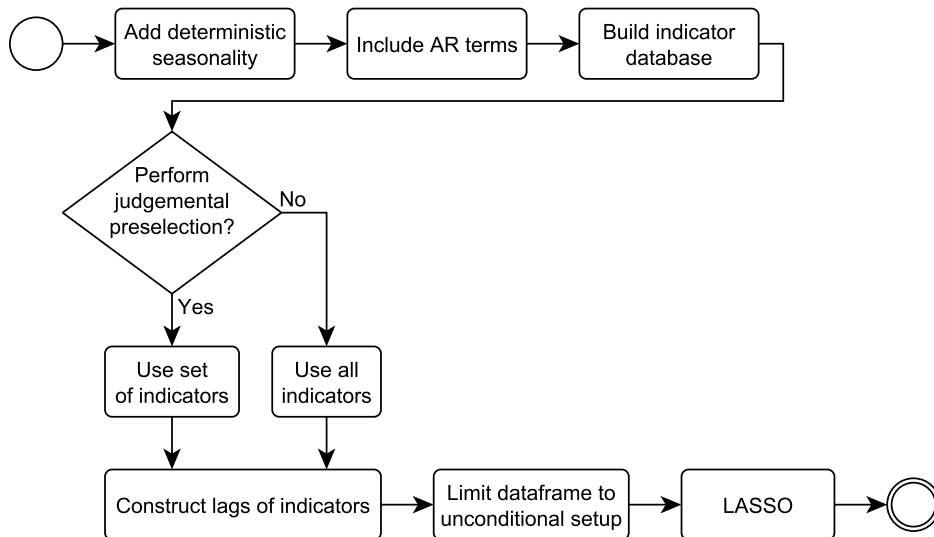


Figure 1: Methodology flowchart.

3.1.1. Seasonality information

In business forecasting seasonality is often an important source of variability. Given the typically limited sample size of sales data, it is very challenging to distinguish between stochastic and deterministic seasonality. From a modelling perspective deterministic seasonality is modelled via $(S - 1)$ seasonal binary dummies, while stochastic seasonality can be modelled via two different approaches. The first approach, appropriate when there is no seasonal

unit root, includes the original variable with a seasonal lag, which results in losing S data point from the training sample. The second approach takes the seasonal difference of the time series $y_t - y_{t-s}$, and then models the differenced series, potentially with additional seasonal lags. This results in losing S data points by the differencing, and potentially another S data points by including seasonal lags of the differenced variable. In this paper we opt to implement deterministic seasonality via seasonal binary dummy variables, to lose as few estimation points as possible, with minimal expected effect on forecasting accuracy, due to the relatively small sample size that is typical in business forecasting (Ghysels and Osborn, 2001).

This will be important when identifying important leading indicators. The seasonal information can be excluded if unnecessary by the LASSO regression, at the variable selection step.

3.1.2. Autoregressive information

We consider additional univariate information by including autoregressive (AR) lags. While macroeconomic leading indicators are typically slow moving, fast moving dynamics can be captured with appropriate AR terms. Furthermore, in contrast to leading indicators, AR terms imply no additional data cost and therefore should be preferred to external variables, if they contain similar information. The lag order is restricted to be smaller than the seasonal lag, since the latter is modelled separately. To simplify the variable selection step we pre-filter the potential AR inputs. This is done by a stepwise search of AR terms with the Akaike Information Criterion corrected for sample size (AICc), in line with the work by Burnham and Anderson (2004). Other information criteria, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) can be used as alternatives (Burnham and Anderson, 2002) and there is no consensus as to which is the best. We conducted a sensitivity analysis and found that AICc performed adequately (see table 4), in line with the findings of Hurvich and Tsai (1991) who compare the use of AICc, AIC and BIC for autoregressive time series modelling, and argue in favour of using AICc. For further details of the stepwise search for AR terms the reader is referred to Hyndman and Khandakar (2008). The identified set of AR inputs is subsequently used as inputs to the LASSO regression, which will evaluate which specific AR terms are included in the final model.

3.1.3. Leading indicators

The last step in building the inputs for LASSO in figure 1, is adding the macroeconomic leading indicators. As the total set of potential inputs can be very large, incorporating expert domain knowledge in the modelling framework could reduce the total number of potential inputs. Since these indicators will be further selected by the LASSO regression the expectation is that experts do not have to perform a very detailed selection, but merely reduce the set to broadly relevant indicators. Alternatively, if domain knowledge is not available, the entire set of leading indicators can be fed to the regression. In our methodology both options are considered, thus permitting us to account the value added by experts.

These indicators are shifted in time, to model any leading dynamics to the target sales variable. The appropriate leading effect of each indicator is unknown. To resolve this we consider lagged versions of the indicators as inputs to the LASSO, which is tasked to identify the most useful lags. This is in contrast to related research with LASSO, where dynamic effects are typically captured in the error term through an Autoregressive Distributed Lag model (Ma et al., 2015). Modelling the lead effect of the indicators simultaneous with the indicators selection, instead of through the error term has two major benefits: (i) the model is more transparent, providing business intelligence to its users, as any lead effects are direct and easy to communicate; (ii) when forecasting, the model tracks substantial changes in the indicators, such as inflection points, timely due to the appropriate lead, and maps them to the forecasts.

The forecasts of the LASSO models are produced by the linear regression:

$$\hat{Y}_{t+h} = \beta_0 + \sum_{s=1}^{S-1} \beta_s D_s + \sum_{r=1}^R \beta_{(S-1+r)} y_{t-r} + \sum_{p=1}^P \beta_{(S-1+R+p)} x_{(t+h),p}, \quad (1)$$

where each of the three sums represents a different type of information, as also shown in figure 1. The first sum over $(S - 1)$ contains information from the seasonal dummies D_s with seasonal length S , the sum over R contains the autoregressive information that is included by lagging the original time series. The last sum over P contains the exogenous inputs considered by the model. Each input is a combination of an indicator and a lag.

3.2. Unconditional forecasts

Forecasts using leading indicators have a practical limitation: only information that is available at the time when the forecast was generated is

available and can be used, as the future values of the leading indicators are unknown. Figure 2 visualises this constraint. When forecasting horizon $h = 1$, all the lagged indicators can be used. As the forecast horizon increases, lower order lags can no longer be used, as they will have no data. For instance, when forecasting horizon $h = 12$, only the indicator that is lagged by 12 months is available and may be used. This setup is referred to as unconditional or ex ante forecasting (Ord and Fildes, 2012). The main difference with conditional forecasting is that the latter assumes that all future values of the indicators are known. Therefore, when forecasting 12 months ahead all low and high order lagged versions of the indicators are assumed to be available.

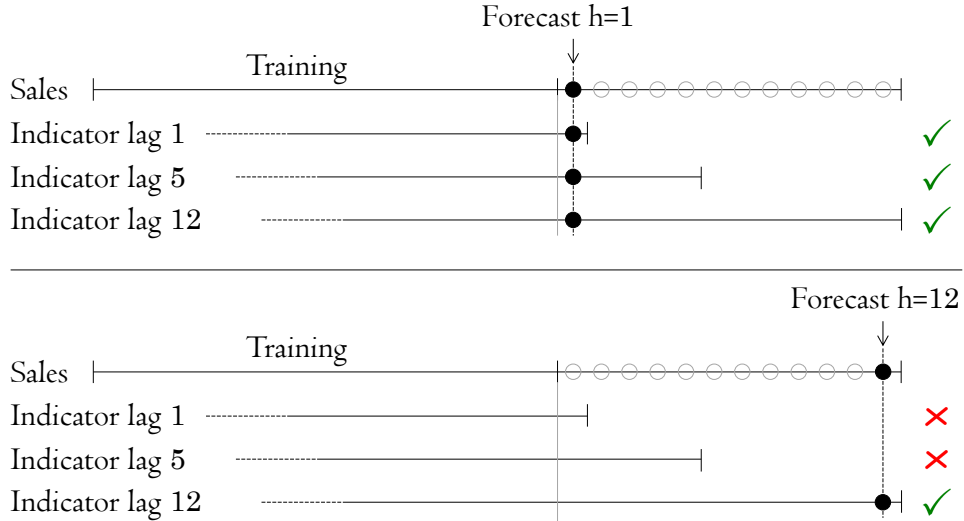


Figure 2: In the unconditional forecasting setup, when forecasting horizon $h=1$, the indicators lag 1 up to lag 12 can be used. However, when forecasting horizon $h=12$, only indicators with lag 12 can be used.

This limited data availability of macroeconomic indicators complicates their use in industry and means that any forecasting model has to be reformulated for each forecast horizon, using only appropriate lagged realisations of the variables. In turn, this limits the available subset of indicators so that only lags bigger or equal to the current forecast horizon are included. The original non-shifted indicators are not used since they are contemporaneous and therefore contain no useful leading information for the forecasts.

This means that for longer horizons only long leads can be captured, as any shorter dynamics are not available to the model, thus limiting the useful information. Note that this limitation is not relevant to univariate lags, and makes the use of the latter important in capturing any shorter-term dynamics.

Furthermore, by constructing unconditional forecasts using only lagged indicators, we avoid predicting all indicators separately. Apart from the apparent computational benefits and not introducing any additional errors to our forecast, the key benefit of this approach is that any hard-to-predict turning points and shocks captured in indicators are used as inputs to our model. If we were to predict the leading indicators this would not be possible.

3.3. Least Absolute Shrinkage and Selection Operator

LASSO regression performs simultaneously coefficient shrinkage and variable selection (Tibshirani, 1996). First, all predictors' values x_{ip} are standardised with zero mean and unit variance. This standardisation ensures that the LASSO does not depend on the units of the predictors (Hastie et al., 2015). The target values y_i are also centred around zero, therefore the intercept β_0 can be omitted in the optimisation procedure. The intercept β_0 is then calculated on the original scale as

$$\beta_0 = \frac{1}{N} \sum_{i=1}^N y_i - \sum_{p=1}^P \beta_p \frac{1}{N} \sum_{i=1}^N x_{ip}, \quad (2)$$

where the estimated β_p of each predictor on the standardised scale remain the same on the original scale. When formulating the model on the original scale, this results in the following linear regression

$$y = \beta_0 + \sum_{p=1}^P \beta_p x_p + \epsilon, \quad (3)$$

where, y is the vector of sales values, x_p is the vector of original values of one indicator, and ϵ are independent and identically distributed normal errors with mean 0 and unknown variance σ^2 . Model parameters are optimised using the following cost function:

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{p=1}^P \beta_p x_{ip} \right)^2 + \lambda \sum_{p=1}^P |\beta_p|. \quad (4)$$

This minimises the residual sum of squares over the training sample n , as ordinary linear regression does, but also penalises it with the sum of the absolute values of the coefficient of potential variables. The second term encourages sparser solutions, as it can result in setting some of the coefficients to zero, when the parameter λ is sufficiently large. This way, LASSO can perform variable selection, potentially retaining only a subset of the variables. The selection of the shrinkage factor λ is critical and is typically determined by using cross-validation.

Tibshirani (1996) argues that a small amount of large effects are better chosen via judgemental selection, but that LASSO performs better when the amount of variables increases, making it ideal for our case. LASSO results in a simple and interpretable model, only containing a subset of the original pool of variables.

Conventional LASSO is computationally intensive, making it difficult to implement on big datasets (Tibshirani, 2011). To overcome this difficulty, Friedman et al. (2010a) developed a more efficient approach for solving generalised linear models with convex penalties such as LASSO. This uses cyclical coordinate descent along the regularisation path, effectively reducing the required computational time. Convergence of each model is obtained faster by fitting a sequence of models with different λ , referred to as a 'warm start'. The reader is referred to Friedman et al. (2010a) for the details of the algorithm.

Since multiple lagged realisations of the input variables are used, multicollinearity may become a modelling issue. When two variables are highly correlated, LASSO will pick one and remove the other, due to shrinkage. Therefore this is an efficient way to deal with multicollinearity, simplifying the modelling process.

4. Empirical Evaluation

4.1. Case Study

Sales data of a multinational company will be used for this analysis. The company is a major supplier of raw material to the global tire industry. The data represents two types of products, supporting passenger cars and trucks. Hereafter, the series are named according to their end market use. Figure 3 shows the series representing the overall sales for the US and the EU with end markets in passenger and truck tires. These series are used in the tactical

sales forecasting for the respective segments. The values on the y-axis are not provided for reasons of confidentiality.

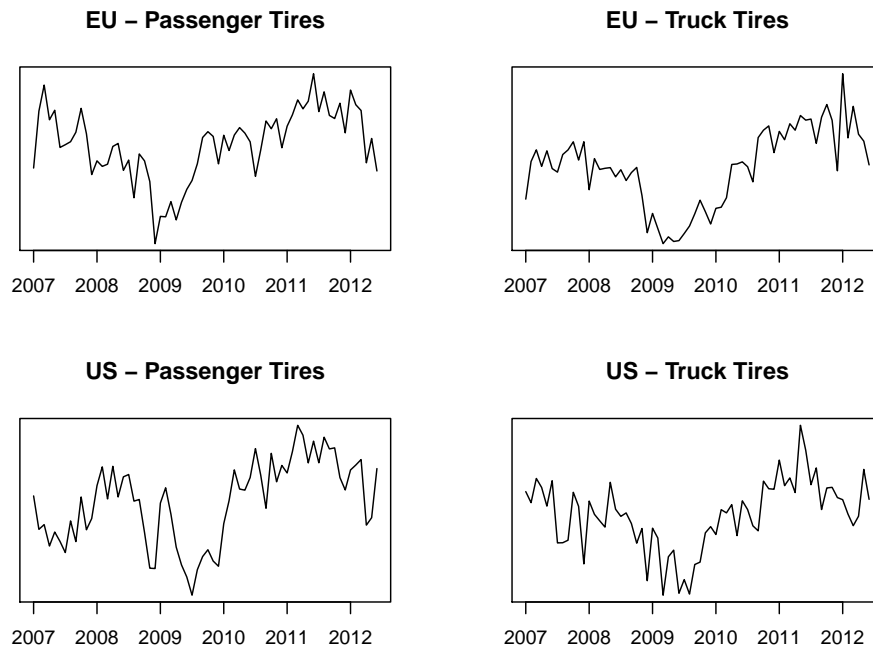


Figure 3: Company sales series and their end markets

In this figure, the effects of the economic crisis in December 2007 to June 2009 are clearly visible in the drop of sales by the end of 2008. Using statistical testing we find that the series exhibit no significant trend, but are seasonal. Note that this testing was done purely for exploration and is not part of the modelling methodology outlined above. As discussed in section 3.1.1, the seasonality information is included in the modelling framework through seasonal dummies, which can be retained or not by LASSO in the variable selection step.

When economic activity increases, more goods are transported by road, and this results in a need for more truck tires. Since the tires are only replaced after they are worn out, the increase in tire replacements will occur some time after the increase in economic activity. The leading effect of the economy, and therefore the leading effect of macroeconomic indicators should have the potential to improve tactical sales forecasts. A similar argument can

be made about passenger tire sales. The variables in our pool of indicators cover a broad spectrum of macroeconomic indicators on a monthly basis. They cover Consumer Price Indexes, labour and earnings measures, financial activities, trade, transportation, manufacturing, retail trade, housing, health care and mining. The macroeconomic indicators originate from the publicly available indicators of the Federal Reserve Economic Data (FRED). Only indicators available on a monthly frequency are considered and incomplete indicators that are no longer updated, or that were only recently introduced, are not considered. This results in a complete set of 67,851 indicators. To provide some insight in the type of macroeconomic indicators considered, a sample of indicators is shown in table A.1, clustered in 8 groups.

After interviewing the company’s supply chain manager the full list was reduced to 1,082 relevant indicators, resulting in a second set of judgementally pre-filtered leading indicators. The discussion in the interview was based on the latest quarterly industry reports of external parties to identify factors that could impact the sales on a tactical level. In the selection, we asked to explicitly state the assumptions why this factor would be important for the company sales. From this, we got a list of keywords which seemed reasonable to the supply chain manager. We selected all the indicators that were related to these keywords in the database. We evaluate the performance of both the judgemental set and the full set, so as to assess the value of the judgement for pre-filtering leading indicators. This follows the suggestion by Lu et al. (2012) that expert opinion can improve the indicator selection. Note that it was not possible for experts in the company to identify a small set of indicators, so as to construct a fully judgementally specified benchmark regression.

4.2. Experimental setup

The in-sample period of all time series covers January 2007 to June 2012, and gradually increases as we perform a rolling origin forecast evaluation, as shown in figure 4. The forecasts for $h = 1$ to $h = 12$ are generated from each new forecast origin. This is done as follows: first, a 12 months forecast is generated starting from July 2012, then the in-sample increases by one month and another 12 months forecast is produced, starting from August 2012. The process is repeated until the complete sample is exhausted. The rolling origin evaluation scheme permits us to collect multiple forecast errors, increasing the reliability of our comparisons and the robustness of the results

to specific forecast origins and test periods. The overall forecast accuracy across forecast origins is summarised for each horizon separately.

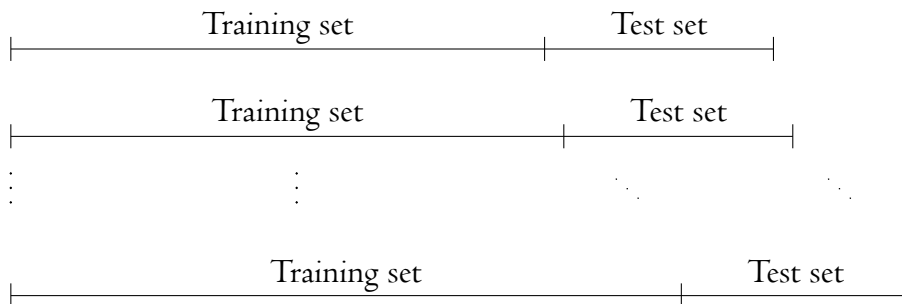


Figure 4: Rolling origin setup, where the initial training set is gradually increased by one period until all the available sample is used.

To measure the accuracy of the forecasts, we report the Mean Absolute Percentage Error (MAPE), which is widely used in practice as it has an intuitive interpretation and is also relevant to the investigated case company:

$$MAPE_h = \frac{1}{n} \sum_{t=1}^n \frac{|Y_{t+h} - \hat{Y}_{t+h}|}{Y_{t+h}}, \quad (5)$$

where Y_{t+h} is the actual sales and \hat{Y}_{t+h} is the forecast for time period $t+h$. The absolute percentage error for each horizon h is aggregated across forecasts from n origins to form the MAPE. Note that we considered various alternative error metrics that provided the same insights and therefore are not reported for brevity.

4.3. LASSO setup

The regularisation parameter λ in Eq. (4) is identified via a 10-fold cross-validation on the in-sample data. To mitigate any potential overfitting due to limited training sample, we select the λ value one standard deviation of the cross-validation errors away from the minimum mean squared cross-validation error (Hastie et al., 2015).

Four variants of LASSO are evaluated. First, the judgementally reduced set of indicators is included in the model to be selected via LASSO. Pre-filtering of the variables simplifies the modelling exercise. These forecasts are

named ‘LASSO (set)’ hereafter. Second, we make autoregressive information available to the LASSO. In contrast to the leading indicators, autoregressive information represents univariate information with no additional external data cost. Therefore we evaluate whether the inclusion of leading indicators still adds value in the presence of appropriate univariate inputs. These forecasts are named ‘LASSO AR (set)’. Third, we use the full set of 67,851 indicators. This is done to evaluate the benefit of judgementally pre-filtering the indicators and assessing the possibility of fully automating the process in the context of tactical forecasting of our case study. These forecasts are named ‘LASSO (all)’. Finally, in the last variant of LASSO, conditional forecasts are produced, where the future values of indicators are assumed to be known. This final variant is named ‘Oracle LASSO (set)’. This last setup provides insight into how much forecast accuracy we lose by not knowing the future values of the exogenous indicators, as discussed in section 3.2. The input variables of the four LASSO variants are lagged appropriately further increasing the dimensionality.

The maximum lag was chosen to be equal to 12, allowing for leading associations up to a year ahead. The lead effect of these indicators on the sales series is not known in advance, so all 12 different lags of the indicators are included in the LASSO model, resulting in a full set of variables (including all possible lags) of 814,212 variables. Since the forecasts are unconditional, we need to restrict the number of available lags for longer forecast horizons. For example, to forecast for $h = 12$ only indicators with lag order 12 are used as inputs. As table 2 illustrates, forecasting longer horizons drastically limits the number of inputs.

Table 2: Number of inputs for horizons 1, 5 and 12.

Forecast horizon	1	5	12
Lags	1-12	5-12	12
Number of variables	814,212	542,808	67,851

For every forecast origin in the experiment, the four models outlined above are refitted for each forecast horizon. The LASSO models are implemented using the ‘glmnet’ package for R (Friedman et al., 2010b).

4.4. Benchmark Models

We benchmark the LASSO models with six univariate models and two conventional regression models that also incorporate leading indicators. The

six extrapolative models considered in this study are Naive, Seasonal Naive, Holt-Winters, Exponential Smoothing family of models (ETS), Auto-Regressive Integrated Moving Average model (ARIMA) and ARIMA without the moving average part (ARI).

The Naive requires no parameter tuning and as such is very simple to implement and use. Any complex forecasting methods should perform better than the Naive to warrant their use. To account for any seasonality, the Seasonal Naive model is also included as a benchmark. ETS has shown good performance in several studies (Hyndman et al., 2002, Gardner, 2006) and is included as a benchmark. ETS is capable of modelling a wide variety of time series, including level, trend or seasonal components that interact in an additive or multiplicative way. This has made it one of the most widely used business forecasting methods in practice. To choose the appropriate exponential smoothing model form (type of error, trend and seasonality) we use AICc. The case company is currently using the Holt-Winters exponential smoothing method to produce the forecasts and therefore we use it as a separate benchmark to evaluate any gains over current practice. ARIMA permits for different model structures to ETS and therefore is used as an additional benchmark. To implement it we use the ARIMA model identification methodology by Hyndman and Khandakar (2008). To fairly assess the usefulness of including purely autoregressive information, which can be readily incorporated in the LASSO models, we build the ARI benchmark forecasts for which the moving average order is restricted to zero. The univariate benchmark models are implemented using the ‘forecast’ package in R (Hyndman, 2017).

These well-known benchmarks rely only on univariate information. Two more benchmarks that incorporate external information are added. The forecasts are formulated using a linear regression model, as in Eq. (1), including a constant, eleven seasonal dummies and leading indicator variables. Based on the judgementally selected subset of indicators, the highest correlated indicator is used for the first model: ‘Lin Reg’. The second model, ‘Step Reg’, uses forward stepwise regression to select from a pool of the 20 highest correlated variables of the judgemental subset of indicators. We are not using the complete set of indicators in ‘Step Reg’ because of the computational cost of the required calculations that make the selection practically impossible. Furthermore, we produce conditional forecasts for the regression benchmarks, which are distinguished with the prefix ‘Oracle’.

5. Results

The forecast accuracy results are provided in table 3. For each method, the MAPE from 1 to 12 steps ahead is provided, followed by the overall accuracy in the last column. The values in boldface highlight the best performing model for each forecast horizon. Overall, the ‘LASSO (set)’ model performs best.

In table 3 we highlight with letters (a-e) different interesting points that are discussed. (a) First, the company benchmark Holt-Winters can already be improved by implementing the full ETS, resulting in a 17.7% reduction in overall error. Although Holt-Winters is performing best on short horizons, ETS outperforms all the univariate models from forecast horizon $h = 4$ up to $h = 12$.

Considering forecasts that use leading indicator information (b) both linear regression benchmarks perform poorly. In particular for ‘Step Reg’ this is attributed to two issues. First, multicollinearity among the different included leading indicators causes a weaker performance. Second, the greedy search strategy of ‘Step Reg’ explores only a limited subset of variables and may eventually ignore useful variables, or remove them (Hastie et al., 2011). Further tests with Bayesian Information Criteria (BIC) for the ‘Step Reg’ have been performed. Although the results exhibit improved performance, this does not change the conclusion of table 3. Hence, these results are omitted for brevity. (c) Comparing ‘LASSO (all)’ and ‘LASSO (set)’ we find that the expert selection improves the overall forecast accuracy. For different horizons, the judgemental selection helps LASSO to retain more relevant variables. Although the performance of ‘LASSO (all)’ is worse than ‘LASSO (set)’, our results suggest that it is a viable alternative when judgementally pre-filtering indicators is not possible, with ‘LASSO (all)’ performing substantially better than the regression benchmarks. The difference in accuracy can help gauge the trade-off between performing a judgemental selection or accepting the lower accuracy in terms of cost. (d) When additional univariate information from the auto-regressive process is included into the ‘LASSO (set)’ model, the MAPE improves on short term, but not for longer horizons, where any autoregressive inputs are based on shorter-horizon forecasted values. This observation reflects the limited benefits of autoregressive information that were observed for the benchmark methods, in particular given the volatile market that the case time series describe.

Note that although a moving average process is not included into the

Table 3: MAPE for unconditional forecasting setup

MAPE for Horizon h:	1	2	3	4	5	6	7	8	9	10	11	12	Overall
Naive	16.3	16.5	18.7	17.5	16.9	18.0	12.0	15.2	17.8	20.3	21.0	16.2	17.2
Seasonal Naive	28.1	25.7	24.4	23.4	21.7	19.7	19.3	19.4	18.6	19.2	18.7	16.2	21.2
Holt-Winters	11.0	13.8	13.7	16.1	16.4	17.3	18.5	19.7	21.8	25.2	25.8	24.0	18.6 ^a
Exponential Smoothing (ETS)	13.5	15.8	15.9	15.7	15.0	15.3	10.4	13.1	16.7	18.5	18.1	16.0	15.3 ^a
ARIMA	15.3	17.4	16.2	16.8	15.8	15.4	15.2	15.4	17.5	20.7	22.6	21.4	17.5 ^e
ARI	16.4	18.5	17.3	17.5	18.4	18.5	17.8	17.3	19.2	23.1	25.0	24.8	19.5 ^e
Lin Reg	15.7	17.1	17.2	17.9	15.4	16.5	18.1	16.2	19.4	29.0	30.7	30.3	20.3 ^b
Step Reg	17.9	18.0	13.4	16.6	16.2	20.2	21.3	22.4	22.3	25.0	26.5	33.2	21.1 ^b
LASSO (all)	13.3	13.6	13.5	13.8	13.7	13.9	14.2	16.5	17.5	17.9	20.7	18.2	15.6 ^c
LASSO (set)	13.5 ^d	13.2^d	12.5	13.5	13.4	13.4	11.5	12.8	15.4	19.2	20.6	22.5	15.1^c
LASSO AR (set)	13.0 ^d	13.4 ^d	12.6	13.4	13.2	13.2	12.4	13.1	16.2	20.8	24.9	22.5	15.7

Letters (a-e) highlight different interesting points that are discussed in the text.

LASSO model, we can deduce the potential of this univariate information by comparing ‘ARIMA’ and ‘ARI’ in (e), where the former improves accuracy over the latter by 10.3%.

The findings in table 3 hold when Mean Absolute Scaled Error (MASE) or Geometric Mean Relative Absolute Error (GMRAE) were used.

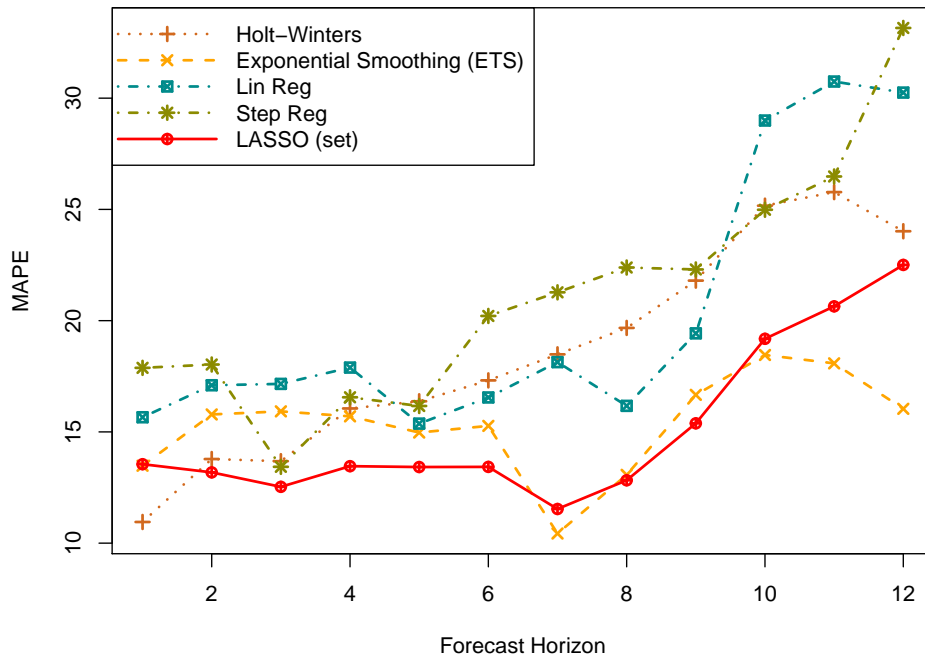


Figure 5: Forecast accuracy over the different forecast horizons.

To compare the results over different horizons, figure 5 displays the MAPE for each forecast horizon of four benchmark models and the ‘LASSO (set)’ model. Naturally long term forecasts are on average associated with higher errors.

Comparing the different benchmark models, we see that ETS in general outperforms all other univariate benchmarks methods consistently. We note an unexpected good performance of ETS and LASSO (set) for horizons $h = 7$ and $h = 8$. This is due to one time series, who exhibits a particular good forecast accuracy for $h = 7$ over 7 rolling origins. To a lesser extent this is

also true for $h = 8$. On short to mid-term, we can note a substantial difference between LASSO and the benchmarks. On longer forecast horizons, we see that ETS gains over LASSO. For these longer forecast horizons only higher orders of indicator lags are included in the model, due to the unconditional forecasting setup (see figure 2), as indicated in table 2, which do not exhibit as strong predictive information compared to shorter leads. In other words, it is easier to find a quarterly leading macroeconomic indicator than an indicator that is leading one year ahead. This is indicative of the practical limitations of causal models with exogenous information for long term forecasting. However, it is important to note that a major difference between ‘LASSO (set)’ and the various benchmarks is that the former provides insight on which indicators are important and how they affect the sales.

In addition to (d) in table 3, we provide a sensitivity analysis for the choice of information criterion for the AR pre-filtering in table 4. We compare AICc, AIC and BIC. We find BIC to be best overall, followed by AIC and AICc that exhibit identical performance, but none is consistently best for all forecast horizons. For short term forecasting AIC and AICc perform better. On the other hand, BIC provides better results on longer term forecasts. BIC penalises additional parameters more strongly, as a result it includes less terms. Nonetheless, we find that LASSO (set) is still superior to either LASSO AR (set) BIC or LASSO AR (set) AICc.

All models that use exogenous information, depend highly on the quality of this information and obviously on the availability of the exogenous inputs for the future periods. To quantify this we provide the accuracy performance when we assume that the future values of the indicators are known, i.e. for the ‘Oracle’ forecasts. Table 5 provides the conditional forecasting results. Note that the ‘Oracle’ models can potentially use short order lags for the input variables even for long forecast horizons, in contrast to the unconditional forecasts, as discussed above.

The ‘Oracle LASSO (set)’ results show that the overall forecast performance could improve by 8.6% compared to ‘LASSO (set)’ if the values of the indicators in the future were known. The MAPE of the conditional model is lower on each horizon. This indicates that shorter-term leading information is relevant, which is not possible to retain in longer term unconditional forecasts. Observe that although the results indicate that if reliably predicted indicators were available these can increase the forecasting performance, the loss of accuracy in our unconditional forecasting setup is smaller than the observed gains over the current forecasting for the case company.

Table 4: MAPE for different information criteria in the autoregressive pre-filter for LASSO

MAPE for horizon h:	1	2	3	4	5	6	7	8	9	10	11	12	Overall
LASSO (set)	13.5	13.2	12.5	13.5	13.4	13.4	11.5	12.8	15.4	19.2	20.6	22.5	15.1
LASSO AR (set) AICc	13.0	13.4	12.6	13.4	13.2	13.2	12.4	13.1	16.2	20.8	24.9	22.5	15.7
LASSO AR (set) AIC	13.0	13.4	12.6	13.4	13.2	13.2	12.4	13.1	16.2	20.8	24.9	22.5	15.7
LASSO AR (set) BIC	13.6	13.5	13.0	13.5	13.1	13.3	11.7	13.0	15.9	20.4	22.1	23.2	15.5

Table 5: MAPE for conditional forecasting setup

MAPE for Horizon h:	1	2	3	4	5	6	7	8	9	10	11	12	Overall
Oracle Lin Reg	15.9	17.2	18.0	20.5	20.9	23.3	22.6	25.8	28.8	29.2	30.2	30.0	23.5
Oracle Step Reg	12.5	15.2	14.6	14.6	14.7	15.1	17.7	16.4	18.1	22.7	24.0	25.1	17.6
Oracle LASSO (set)	13.1	12.6	12.1	12.9	13.0	12.8	11.8	12.7	14.3	15.7	16.1	18.3	13.8

In contrast to the results in table 3, ‘Oracle Step Reg’ improves upon the regression that includes only one indicator ‘Oracle Lin Reg’. The conditional stepwise regression can benefit from better input information, while ‘Oracle Lin Reg’ suffers from overfitting, performing worse than ‘Lin Reg’.

Next, we turn our attention to the inputs that are used by LASSO. It selects on average 15 variables for the first horizon $h = 1$, but only 10 for $h = 12$. From these selected variables, as we build different models for each forecast horizon, several lagged indicators are retained throughout. From the 12,984 initial variables considered by ‘LASSO (set)’, it ends up selecting a much smaller pool of 88 variables on average, for each time series across all horizons and origins. In this pool, on average 61 indicators are unique and the remaining are reused.

For the different time series, one to seven lagged indicators are selected by the LASSO all the time across origins and horizons. This set of variables naturally decreases across longer horizons, as our forecasts are unconditional. This indicates that these indicators contain important information, as they are picked consistently. These indicators have a stronger link with the sales series and represent different types of information, such as employment in automobile dealers, national passenger car registrations and Consumer Prices Index for solid fuel prices.

One of the common types of indicators that appears for several time series, over different horizons and rolling origins is the ‘Passenger Car Registrations’. These indicators are available on a country level, so one or multiple of these indicators appear in the model. The leading effect of this type of indicator varies between 1 and 4 months. The selection fits to our initial hypothesis about potential drivers of demand for the case company.

As the horizon increases, we use only longer lags of the indicators. We are interested whether as the horizon increases, if the order of the indicators increases as well or instead different indicators are used. We see that 47 lagged indicators from the pool of 88 variables are selected more than once on different horizons. Seventeen indicators in this pool are selected with different lags over multiple horizons. LASSO selects multiple times the same indicators, but increases the lag of the indicator as the horizon increases. As an individual LASSO model is formulated for each separate horizon, 11 of these 17 indicators have simultaneously multiple lags in several of the individual LASSO models.

Based on these findings we argue that LASSO under the proposed framework was able to identify and select important leading indicators, with the

resulting improvements in forecasting accuracy over current practice and benchmarks.

6. Conclusion

In this paper, we propose a framework to improve tactical sales forecasting using macroeconomic leading indicators. Tactical sales forecasts span typically up to 12 months ahead. In this time scale the changing characteristics of the economy can impact sales significantly. We construct forecasts under the hypothesis that exogenous macroeconomic information can improve accuracy and provide insight in the relevant market dynamics.

The contributions of our work can be summarised as follows: (i) we demonstrate the usefulness of macroeconomic leading indicators for tactical sales forecasting; (ii) propose a fully automatic methodology that is able to select appropriate indicators and their lead order from massive set; and (iii) demonstrate the benefits of incorporating expert knowledge for pre-filtering relevant indicators. We also conduct a comparison of unconditional (ex ante) forecasts against conditional (ex post) ones, quantifying the performance loss of regression models in practical settings that the future values of regressors are typically unknown.

Our findings indicate that (i) the proposed methodology can improve accuracy over standard practice and established statistical benchmarks, while (ii) providing market insights to managers of relevant leading indicators and their effect on sales. We also find that (iii) managerial judgement is useful in pre-filtering group of potentially useful indicators and quantify the gains, providing insight into the potential accuracy trade-off between using a fully statistical approach or experts; each alternative implying a different cost for a company.

For the case study company we find accuracy improvements over current practice by 18.8%. In interviews with the global supply chain manager the expectation is that this increase in accuracy results in substantial reductions of Work-In-Process and Work-In-Capital. The global supply chain manager of this company argued that this difference is sufficient to physically relocate constructed machine resources to different production plants to take advantage of the increased forecast accuracy.

Furthermore, given the provided insights on key indicators and their effects, the company can be more agile to potential opportunities and threats coming from the economic environment. The model allows management to

simulate what-if scenarios of substantial changes in the macroeconomic conditions, for instance stemming from wider economic events or political decisions. Since the important leading indicators are known, as well as their lead order, wider effects can be accounted to these and in turn to the company sales.

In this analysis we considered only monthly macroeconomic indicators. Several important macroeconomic indicators are available on different frequencies, such as on a quarterly or yearly basis. Inclusion of different frequency variables can potentially augment the available information in the model, as well as allow us to use a wider set of inputs beyond macroeconomic indicators, such as information on prices, promotional activities or competitive actions. In our case this latter information was not available. A further modelling aspect that was not investigated is the inclusion of potential non-linear effects from the various indicators. As no prior knowledge is given of existing non-linear relationships between the sales and the set of indicators, and the volume of indicators to explore is massive, this increases the complexity of the problem. How to best achieve these in a tactical forecasting setting is an interesting open research question.

Appendix A. Table of example FRED indicators

Table A.1: Table of example FRED indicators

Category	Type	Indicator Name	Unit	Freq.	Transformation ^a
Consumption	Beverages	Consumer Price Index for All Urban Consumers: Alcoholic beverages	Index 1982=1984=100	M	SA
	Beverages	Merchant Wholesalers Except Manufacturers' Sales Branches and Offices Sales: Nondurable Goods: Beer Wine and Distilled Alcoholic Beverages Sales	Millions of Dollars	M	NSA
	CPI	Consumer Price Index for All Urban Consumers: All Items	Index 1982=1984=100	M	SA
	CPI	Consumer Price Index for All Urban Consumers: New vehicles	Index 1982=1984=100	M	NSA
	CPI	Harmonized Index of Consumer Prices: All Items for Euro area (17 countries)	Index 2005=100	M	NSA
	Durable goods	Personal Consumption Expenditures: Durable Goods	Billions of Dollars	M	SAAR
	Food	Real Retail and Food Services Sales	Millions of Dollars	M	SA
	Harmonized	Harmonized Index of Consumer Prices: All Items for United States	Index 2015=100	M	NSA
	Hospitality/Leisure	All Employees: Leisure and Hospitality	Thousands of Persons	M	SA
	Nondurable goods	Personal Consumption Expenditures: Nondurable Goods	Billions of Dollars	M	SAAR
	Sentiment	University of Michigan: Consumer Sentiment	Index 1966:Q1=100	M	NSA
	Services	Personal Consumption Expenditures: Services	Billions of Dollars	M	SAAR
	Feedstock	Commodities	Consumer Price Index - Services - Total for United States	Index 1957=1959=100	M
Commodities		Crude Oil Prices: Brent - Europe	Dollars per Barrel	M	NSA
Commodities		US Regular All Formulations Gas Price	Dollars per Gallon	M	EOP - NSA
Metals		Producer Price Index by Commodity for Metals and Metal Products: Iron and Steel	Index 1982=100	M	NSA
Financial	Financial	Kansas City Financial Stress Index	Index	M	NSA
	Financial	Consumer Revolving Credit Owned by Finance Companies - Outstanding	Billions of Dollars	M	EOP - NSA
	Government	Interest Rates - Government Securities - Treasury Bills for United States	Percent per Annum	M	NSA
	Industry	Commercial and Industrial Loans - All Commercial Banks	Billions of U.S. Dollars	M	SA
	Inflation	5-Year Forward Inflation Expectation Rate	Percent	M	NSA
	Insurance	Personal current transfer receipts: Government social benefits to persons: Unemployment insurance	Billions of Dollars	M	SAAR
	Retail	Retail Money Funds	Billions of Dollars	M	SA
	Building	New Private Housing Units Authorized by Building Permits	Thousands of Units	M	SAAR
	Housing	Housing Starts: Total: New Privately Owned Housing Units Started	Thousands of Units	M	SAAR
	Housing Indexes	S&P/Case-Shiller 20-City Composite Home Price Index S&P/Case-Shiller U.S. National Home Price Index	Index Jan 2000=100 Index Jan 2000=100	M M	SA NSA
Import/Export	Export	Export (End Use): All exports excluding food and fuels	Index Dec 2010=100	M	NSA
	Goods	U.S. Exports of Goods to Japan - F.a.s. basis	Millions of Dollars	M	NSA
	Machinery	Import (End Use): Oil drilling - mining and construction machinery and equipment	Index 2000=100	M	NSA
	Services	Imports of Goods and Services: Balance of Payments Basis	Millions of Dollars	M	SA
	Trade	Trade Balance: Goods and Services - Balance of Payments Basis	Millions of Dollars	M	SA

^aSA=Seasonally Adjusted, NSA=Not Seasonally Adjusted, SAAR=Seasonally Adjusted Annual Rate, EOP=End of Period

Industrial	Business	Total Business: Inventories to Sales Ratio	Ratio	M SA	
	Business	ISM Non-manufacturing: Business Activity Index	Index	M SA	
	Business	Business Tendency Surveys for Manufacturing: Confidence Indicators: Composite Indicators: OECD Indicator for the United States	Normalised (Normal=100)	M SA	
	Construction Equipment	Total Construction Spending: Nonresidential	Millions of Dollars	M SAAR	
	Equipment	Industrial Production: Business Equipment	Index 2012=100	M SA	
	Health	Retail Trade: Building Materials - Garden Equipment and Supplies Dealers	Millions of Dollars	M SA	
	Industry	Total Construction Spending: Health Care	Millions of Dollars	M SAAR	
	Manufacturing	Industrial Production Index G17	Index 2012=100	M SA	
	Manufacturing	Manufacturers' New Orders: Durable Goods	Millions of Dollars	M SA	
	Manufacturing	ISM Manufacturing: New Orders Index	Index	M SA	
	Mining	Capacity Utilization: Manufacturing (NAICS)	Percent of Capacity	M SA	
	PPI	Industrial Production: Mining: Drilling oil and gas wells	Index 2012=100	M NSA	
	PPI	Producer Price Index for All Commodities	Index 1982=100	M NSA	
	Utilities	Producer Price Index by Commodity for Final Demand: Finished Goods	Index 1982=100	M NSA	
	Wholesale	Industrial Production: Electric and Gas Utilities	Index 2012=100	M SA	
		Merchant Wholesalers: Inventories to Sales Ratio	Ratio	M SA	
	Labour	Earnings	Average Hourly Earnings of Production and Nonsupervisory Employees: Retail Trade	Dollars per Hour	M NSA
		Education	Unemployment Rate: College Graduates: Bachelor's Degree and Higher - 25 years and over	Percent	M NSA
		Employment	All Employees: Total Nonfarm Payrolls	Thousands of Persons	M SA
		Employment	Civilian Employment-Population Ratio	Percent	M SA
Employment		Average Weekly Hours of All Employees: Total Private	Hours	M SA	
Health		Job Openings: Education and Health Services	Level in Thousands	M SA	
Hours		Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	Hours	M SA	
Household survey		Civilian Unemployment Rate	Percent	M SA	
Labour force		Civilian Labour Force Participation Rate	Percent	M SA	
Services		All Employees: Professional and Business Services: Temporary Help Services	Thousands of Persons	M SA	
Unemployment		Civilian Unemployment Rate US	Percent	M SA	
Unemployment		Unemployment Level US	Thousands of Persons	M SA	
Transportation		Transportation	Producer Price Index by Commodity for Transportation Equipment: Motor Vehicles Parts	Index 1982=100	M NSA
		Transportation	Freight Transportation Services Index	Chain-type Index 2000=100	M SA
		Transportation	Total Vehicle Sales	Millions of Units	M SAAR
		Transportation	Motor Vehicle Retail Sales: Domestic Autos	Thousands of Units	M SA
		Vehicles	Light Weight Vehicle Sales: Autos and Light Trucks	Millions of Units	M SAAR
		Vehicles	Moving 12-Month Total Vehicle Miles Travelled	Millions of Miles	M NSA
		Vehicles	Motor Vehicle Retail Sales: Heavy Weight Trucks	Millions of Units	M SAAR

Acknowledgements

This research is funded by IWT-Vlaanderen (project IWT 130843) and Solventure NV. We would like to thank the anonymous reviewers for their insightful comments that helped us substantially improve the paper.

References

- Ali, Ö. G., Sayn, S., van Woensel, T., and Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10):12340 – 12348.
- Aviv, Y. (2001). The effect of collaborative forecasting on supply chain performance. *Management Science*, 47(10):1326 – 1343.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135 – 171.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191 – 221.
- Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133 – 1150.
- 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.
- Bertrand, J.-L., Brusset, X., and Fortin, M. (2015). Assessing and hedging the cost of unseasonal weather: Case of the apparel sector. *European Journal of Operational Research*, 244(1):261 – 276.
- Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169 – 194. Common Features.
- Bulligan, G., Marcellino, M., and Venditti, F. (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting*, 31(1):188 – 206.
- Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer, New York.

- Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.
- Currie, C. S. and Rowley, I. T. (2010). Consumer behaviour and sales forecast accuracy: What’s going on and how should revenue managers respond & quest. *Journal of Revenue & Pricing Management*, 9(4):374 – 376.
- Dalrymple, D. J. (1987). Sales forecasting practices: Results from a united states survey. *International Journal of Forecasting*, 3(34):379 – 391. Special Issue on: Forecasting in Marketing.
- Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2):357 – 373. Dynamic Econometric Modeling and Forecasting.
- 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.
- Forni, M. and Reichlin, L. (1996). Dynamic common factors in large cross-sections. In Durlauf, S., Helliwell, J., and Raj, B., editors, *Long-Run Economic Growth*, Studies in Empirical Economics, pages 27 – 42. Physica-Verlag HD.
- Forni, M. and Reichlin, L. (1998). Let’s get real: A factor analytical approach to disaggregated business cycle dynamics. *The Review of Economic Studies*, 65(3):453 – 473.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010a). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.
- Friedman, J. H., Hastie, T., and Tibshirani, R. (2010b). *glmnet: lasso and elastic-net regularized generalized linear models*. R package.
- Gardner, E. S. (2006). Exponential smoothing: The state of the art - part II. *International journal of forecasting*, 22(4):637 – 666.

- Ghysels, E. and Osborn, D. R. (2001). *The econometric analysis of seasonal time series*. Cambridge University Press.
- Guo, Z., Wong, W., and Li, M. (2013). A multivariate intelligent decision-making model for retail sales forecasting. *Decision Support Systems*, 55(1):247 – 255.
- Hastie, T., Tibshirani, R., and Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2nd edition.
- Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC Press.
- Haupt, H., Kagerer, K., and Steiner, W. J. (2014). Smooth quantile-based modeling of brand sales, price and promotional effects from retail scanner panels. *Journal of Applied Econometrics*, 29(6):1007 – 1028.
- Huang, T., Fildes, R., and Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2):738 – 748.
- Hurvich, C. M. and Tsai, C.-L. (1991). Bias of the corrected aic criterion for underfitted regression and time series models. *Biometrika*, 78(3):499 – 509.
- Hyndman, R. J. (2017). *forecast: Forecasting functions for time series and linear models*. R package version 8.0.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3):1 – 22.
- 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.
- Klassen, R. D. and Flores, B. E. (2001). Forecasting practices of canadian firms: Survey results and comparisons. *International Journal of Production Economics*, 70(2):163 – 174.

- Kourentzes, N. and Petropoulos, F. (2015). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*.
- Lang, S., Steiner, W. J., Weber, A., and Wechselberger, P. (2015). Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. *European Journal of Operational Research*, 246(1):232 – 241.
- Lawrence, M., O’Connor, M., and Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research*, 122(1):151 – 160.
- Leitner, J. and Leopold-Wildburger, U. (2011). Experiments on forecasting behavior with several sources of information – a review of the literature. *European Journal of Operational Research*, 213(3):459 – 469.
- Li, J. and Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, 30(4):996 – 1015.
- Lu, C.-J., Lee, T.-S., and Lian, C.-M. (2012). Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. *Decision Support Systems*, 54(1):584 – 596.
- Ma, S., Fildes, R., and Huang, T. (2015). 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*.
- Mol, C. D., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318 – 328. Honoring the research contributions of Charles R. Nelson.
- Naser, H. (2015). Analysing the long-run relationship among oil market, nuclear energy consumption, and economic growth: An evidence from emerging economies. *Energy*.

- Ng, S. T., Skitmore, M., and Wong, K. F. (2008). Using genetic algorithms and linear regression analysis for private housing demand forecast. *Building and Environment*, 43(6):1171 – 1184.
- Ord, K. and Fildes, R. (2012). *Principles of business forecasting*. Cengage Learning.
- Petropoulos, F., Fildes, R., and Goodwin, P. (2015). Do big losses in judgmental adjustments to statistical forecasts affect experts behaviour? *European Journal of Operational Research*.
- Sanders, N. R. and Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, 31(6):511 – 522.
- Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. *NBER Working Paper*.
- Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167 – 1179.
- Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147 – 162.
- Stock, J. H. and Watson, M. W. (2006). Chapter 10 forecasting with many predictors. volume 1 of *Handbook of Economic Forecasting*, pages 515 – 554. Elsevier.
- 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.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267 – 288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273 – 282.

- Trapero, J. R., Kourentzes, N., and Fildes, R. (2012). Impact of information exchange on supplier forecasting performance. *Omega*, 40(6):738 – 747. Special Issue on Forecasting in Management Science.
- Trapero, J. R., Kourentzes, N., and Fildes, R. (2014). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, 66(2):299 – 307.
- Trapero, J. R., Pedregal, D. J., Fildes, R., and Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, 29(2):234 – 243.
- Weber, A., Steiner, W. J., and Lang, S. (2017). A comparison of semiparametric and heterogeneous store sales models for optimal category pricing. *OR Spectrum*, 39(2):403–445.
- Weller, M. and Crone, S. F. (2012). Supply chain forecasting: Best practices & benchmarking study.
- Williams, B. D., Waller, M. A., Ahire, S., and Ferrier, G. D. (2014). Predicting retailer orders with POS and order data: The inventory balance effect. *European Journal of Operational Research*, 232(3):593 – 600.
- Yap, G. and Allen, D. (2011). Investigating other leading indicators influencing australian domestic tourism demand. *Mathematics and Computers in Simulation*, 81(7):1365 – 1374. Selected Papers of the Combined IMACS World Congress and MSSANZ 18th Biennial Conference on Modelling and Simulation, Cairns, Australia, 13-17 July, 2009.