Impact of Information Exchange on Supplier Forecasting Performance

Juan R. Trapero*, N. Kourentzes, R. Fildes

Lancaster University
Department of Management Science, LA1 4YX, UK

Abstract

Forecasts of demand are crucial to drive supply chains and enterprise resource planning systems. Usually, well-known univariate methods that work automatically such as exponential smoothing are employed to accomplish such forecasts. The traditional Supply Chain relies on a decentralised system where each member feeds its own Forecasting Support System (FSS) with incoming orders from direct customers. Nevertheless, other collaboration schemes are also possible, for instance, the Information Exchange framework allows demand information to be shared between the supplier and the retailer. Current theoretical models have shown the limited circumstances where retailer information is valuable to the supplier. However, there has been very little empirical work carried out. Considering a serially linked two-level supply chain, this work assesses the role of sharing market sales information obtained by the retailer on the supplier forecasting accuracy. Weekly data from a manufacturer and a major UK grocery retailer have been analysed to show the circumstances where information sharing leads to improved forecasting accuracy. Without resorting to unrealistic assumptions, we find significant evidence of benefits through information sharing with substantial improvements in forecast accuracy.

Keywords: Bullwhip effect, Supply chain, Supply chain collaboration, Bullwhip ratio, Neural Networks

*Corresponding author.
Email addresses: j.traperoarenas@lancaster.ac.uk (Juan R. Trapero), n.kourentzes@lancaster.ac.uk (N. Kourentzes), r.fildes@lancaster.ac.uk (R. Fildes)

Preprint submitted to Omega *August 1, 2011*

1. Introduction and Background

Since the beginning of the 20th century, one of the main problems that Supply Chain Management has had to face is the bullwhip effect [1]. The phenomenon consists of demand variability amplification when moving upwards in the supply chain [2]. Among the consequences of this amplification, for instance, we might find excess inventory, poor customer service and poor product forecasts. Demand signal processing, rationing gaming, order batching, and price variations are the four main sources of the bullwhip effect sources identified by Lee et al. [3].

In order to avoid the bullwhip effect some authors suggest supply chain collaboration as a mean to ameliorate it, see [4] and references herein. The idea behind supply chain collaboration is to find a global optimal solution for all supply chain members instead of different sub-optimal solutions for each one [5]. Information sharing is a way to accomplish such collaboration [6], [7], [8]. In fact, information transparency based on sharing customer demand as well as other variables such as inventories and orders is one of the ten principles proposed in [1] to achieve bullwhip reduction. However, there has been little empirical evidence on the benefits of collaboration as it affects forecasting accuracy and therefore supply chain demand variability amplification. The aim of this paper is to suggest models for incorporating downstream sales information and provide evidence of the improved accuracy that such information sharing can achieve.

Holweg et al. in [7] suggest a classification of four different supply chain collaborations depending on the extent of the collaboration in planning and inventory control. According to that scheme, we may find: i) The traditional supply chain, where no collaboration is established; ii) Information exchange, the supplier and retailer agree a planning collaboration; iii) Vendor Managed Replenishment, here supply chain members collaborate in terms of inventory; and iv) Synchronised supply, where an integrated planning and inventory collaboration is put in place. In this paper, we will be focused on analyzing the benefits of planning collaboration, thus, only the first two types will be considered.

Among the benefits of a planning collaboration, an improvement in forecasting accuracy is expected by reducing uncertainty and removing delays in translating the demand signal [7]. However, there is no general agreement in the literature at this point. In fact, some authors, based on analytical models claim that the information available in the market retail sales can be translated up the stages of a supply chain based only on the orders received [9]. Thus, a chain echelon might retrieve such information by means of an appropriate filter of the downstream information. Here, the term “filter” refers to the mathematical equation that is used to model the demand structure in order to separate valuable information from noise [10]. For instance, assuming that the customer demand can be described by an Autoregressive Integrated Moving Average (ARIMA) filter [11], in particular, a first-order Autoregressive process AR(1) and the inventory management follows an Order-Up-To policy, the resultant order signal is a first order Autoregressive-first order Moving Average process ARMA(1,1). Thus, the second echelon player can obtain the available information on the market sales by filtering the order signal with an ARMA(1,1) structure, and consequently, avoiding the necessity of investing in interorganisational systems for information sharing [9], [12], [13], [14].

However, in order to make the problem mathematically tractable those works rely on restrictive assumptions that tend to be highly constrained versions of reality [15]. For instance, they do not include the influence of promotions, price reductions and advertisements even when price variation is one of the four causes that generate demand amplification [3]. In addition, since those factors are difficult to include in statistical models, they are usually introduced into the forecasting process by managers that judgmentally adjust forecasts provided by a Forecasting Support System, [16], [17]. In the work led by Fildes and Goodwin in [18] a survey of 149 forecasters was conducted. It was found that, on average, 75 percent of the forecasts involved management judgment. In fact, according to the company-based research in [16] and [19] adjustments were common and may be made in between 60 and 90 percent of forecasts, respectively. Therefore, insights obtained from theoretical developments that do not include the aforementioned circumstances can be limited [15].

On the other hand, some authors based on real customer demand datasets, state that information sharing improves forecasting accuracy [15], [20], [12]. Byrne and Heavey in [15] considered a complex supply chain structure with multiple customers, distributors and product families. They concluded that information sharing could lead to potential total supply chain costs savings of up to 9.7%. They chose as forecasting techniques Single Moving Average and Double Exponential Smoothing. Kelepouris et al. in [20] analyzed the impact of information sharing on the bullwhip effect on the basis of a simulated two-stage supply chain, between a warehouse and store of the same company, using real demand data. In this particular case, forecasts were computed using single and double exponential smoothing and information sharing was captured by substituting the orders observed by the warehouse with those observed at store level. The results concluded that information sharing resulted in 21% warehouse order variability reduction. In [12], an attempt to minimize the gap between theory and practice was done by analyzing the influence of supply chain collaboration employing real Electronic Point of Sales (EPOS) data and the orders generated by the retailer in a two-stage supply chain. Three SKUs with different level of sales and without promotions were considered. Assuming that real customer demands follow an AR(1) structure, they concluded that sharing EPOS data could reduce the supplier's Standard Deviation of the Prediction Errors by between 8% and 19% and that the noise element contained in the EPOS data might be the major source of the information sharing benefit. However, given the small sample and the restrictive assumptions, more empirical work is required to extend those results.

In general, works focused on theoretical analyses assume that the demand structure is known, usually represented as an ARIMA model [12]. Nonetheless, the identification process might not be a trivial step, particularly, when price variations are introduced in the process via judgmental adjustments and there is a high number of SKUS as happens in the grocery retailer industry. As a consequence, empirical work ([5], [26]) show that companies use simpler methods such as single moving average and exponential smoothing techniques. In addition, a question remains open: How information sharing affects forecasting methods selection? For instance, in [22] it is shown that the supplier's demand structure in case no information is shared is an ARMA(1,1), however, if information sharing is enabled and market sales is used instead of retailer's orders, its demand structure follows an ARX model, where the market sales information is the exogenous variable. Nonetheless, case studies presented in [5] and [26] use the same forecasting techniques regardless of whether the information is exchanged. In this sense, as Lapide in [21] has remarked, the problem remains of integrating market sales information in forecasting processes. In other words, the fact of sharing information does not imply that there is an effective integration of such information into the supplier's planning processes [7].

In this article, automatic system identification procedures are proposed to select the adequate structure for the supplier shipments integrating the retailer sales information. The aim of this methodology is twofold: i) to relax the assumption of a known demand structure; and ii) to propose an automatic technique that permits the retailer's sales information to be integrated into the supplier's planning process. In order to evaluate the benefits of information sharing, real data from a serially linked two-level supply chain was collected. In contrast to previous research, in this study we do not impose any assumptions regarding the data generation process of the demand patterns observed on different levels of the supply chain, thus we model the data using a wide variety of forecasting models in order to identify the most appropriate. Firstly, linear and nonlinear AR models that use the retailer's sales information as an explanatory (exogenous) variable (ARX) are defined to forecast the supplier's demand. Secondly, other univariate techniques such as ARIMA, exponential smoothing, Moving Averages and Neural Networks are employed as benchmarks representing the case where no information is shared between supplier and retailer.

The results show that the supplier can improve its forecasting accuracy substantially by integrating the sales data from the retailer into its forecasting process. Forecasting accuracy is directly connected to inventory management, lower errors result in reduced stock keeping without compromising the service level. In fact, research has shown that forecast errors can increase organization costs from 10% to 30% depending on the organizational environment (see [22]). These results provide evidence that the additional organizational effort and cost to start-up and maintain information transparency in the supply chain may lead to improvements in operations. In addition, this work opens up the door for designing more advanced Forecasting Support Systems able of adding multivariate models among their statistical forecasting techniques in order to incorporate efficiently the retailer's sales information.

The remainder of our article is organized as follows: Section 2 introduces the case study. Section 3 gives a brief description of the models considered in the paper. Section 4 discusses the empirical experiments. Finally, main conclusions are drawn in Section 5.

2. Case study

The supply chain system consists of a serially linked two-level supply chain; see Figure 1. This supply chain consists of a flow of information represented by a dashed line from the market towards the manufacturer and a reverse one regarding materials represented by a solid line, [7]. Market sales and shipments from the manufacturer are the measured variables, indicated by the sensors in Figure 1. There is also a switch that represents the option of sharing information. When the switch is off it means that we are considering the traditional supply chain case, i.e., sales information is not available for the Manufacturer. When it is on, market sales information is available for the manufacturer. Note that in the latter case, the manufacturer has two sources of information: i) the retailer orders; and ii) the market sales information

Data from a manufacturing company specialized in household products has been collected. The data has been sampled weekly between October 2008 and October 2009. This manufacturing company provides products to one of the largest retailers in the UK with a lead-time equal to one week. The data consist of two time series per SKU, the first one corresponds to the shipments received by the retailer from the manufacturer. The second one, is the customer demand measured by the retailer sales.

It should be noted that previous works use the retailer's order data as input to the forecasting methods [9],[12]. Ideally, volume of shipments received by the retailer should be a delayed version of the retailer orders and such an assumption is usually made in simulation models [23]. Nevertheless, volume and delayed orders might not match due to the “backlash” effect, which refers to the resulting impact of the “bullwhip” effect on shipments downstream [24]. In the analysed case study, we use shipments instead of orders because this variable is what managers employ as input to the company’s Forecasting Support System.

The dataset under study comprises 43 Stock Keeping Units (SKU) with 52 observations per SKU. An example is depicted in Figure 2.

2.1. Exploratory Data Analysis

In Figure 2 we can clearly see the demand variance amplification phenomenon. A possible way to measure the bullwhip effect is to use the ratio of the coefficients of variation between the output supplier sales and the input retailer sales [25]. Let the Bullwhip Ratio (BWR) be denoted by:

$$\text{BWR} = \frac{\sigma_{supplier} / \mu_{supplier}}{\sigma_{retailer} / \mu_{retailer}}. \tag{1}$$

where σ_i is the standard deviation and μ_i is the mean for i equal to supplier shipments or retailer sales. Other conventional bullwhip measures use the ratio of the variance (or standard deviations) instead of the coefficient of variation [23]. In this same reference, Dejonckheere et al. propose two other bullwhip measures based on the frequency response plot. However, in order to compute the frequency response plot it is necessary to model the replenishment rule and calculate the corresponding transfer function between the customer demand and retailer orders. For the sake of generality, the replenishment rule used by the retailer is assumed unknown. Thus, in this article we will measure the BWR as defined in (1).

Figure 3 shows the histogram of BWR according to our dataset of 43 SKUs. In this figure we can see that the resulting histogram is bimodal. The first peak is located around BWR=1.5 and the second one is placed at BWR=3.5 and BWR=4 approximately. Note that some SKUs can reach a BWR greater than 4.

Figure 4 plots the relationship between the mean of both the retailer and supplier sales for the SKUs of our database. Given that the relationship is linear with a regression coefficient equal to 1 approximately, we can conclude that the replenishment rule is focused on following the level of the real customer demand. It is interesting to note that usually the BWR is measured by the ratio of standard deviations rather than coefficients of variation by assuming that the means are equal. Figure 4 verifies that in our dataset such an assumption is valid.

However, we can also analyse the relationship between variances instead of means. Figure 5 is a scatter plot between the variance of supplier and retailer sales. In contrast to the previous figure, the linear relationship is not so clear and the regression coefficient is 1.6, that is greater than 1. This observation also illustrates the Bullwhip Effect (BE).

3. Models

Two kind of models have been analysed to find out whether retailer sales information is useful for the supplier to improve its forecasting accuracy. On the one hand, we employ univariate models, such as Single Exponential Smoothing (SES), Autoregressive (AR), Moving Average (MA) and Autoregressive Integrated Moving Average (ARIMA) models, a univariate Neural Network and a Naïve method. These methods only rely on past information of supplier sales to forecast and so, no information sharing is accomplished. We employ both linear and nonlinear methods in order to capture potential nonlinearities in the data and produce adequate benchmarks. In addition, a multivariate ARX model and multivariate Neural Networks have also been used, where suppliers sales and retailer sales are used as dependent and explanatory variables, respectively. We summarize the methods below.

3.1. Naïve and Moving Average

A forecasting method is an algorithm that provides a prediction of the value at a future time period [26]. Many forecasting algorithms are based on the identification of an underlying pattern in a data series, that pattern can be distinguished from randomness by smoothing (averaging) past values. After eliminating that randomness, the algorithm projects the pattern into the future to generate the forecast. A well know method to reduce the time series random variation is the moving average [27]. A moving average forecast of order k, or $MA(k)$, is given by:

$$F_{t+1} = \frac{1}{k} \sum_{t-k+1}^{t} y_i. \tag{2}$$

The order has been identified by minimizing the sum squared error of the one-step-ahead errors. Note that the Naïve approach used in this paper is a MA(1), since the last known data point (y_t) is taken as the forecast for the next period, which is the well known Random Walk.

3.2. Single Exponential Smoothing

Since around 1950 the use of Exponential Smoothing for forecasting has been the most popular forecasting method used in business and industry [28, 29]. Basically, the Single Exponential Smoothing (SES) consists of adjusting the previous forecast by weighting the forecast error, i.e:

$$F_{t+1} = F_t + \alpha(y_t - F_t), \tag{3}$$

where α is a constant between 0 and 1. This parameter may be set on *a priori* grounds that usually is between 0.05 and 0.3, [30]. However, if a reasonable number of observations are available, α can be estimated by minimizing the sum of the one-step-ahead in-sample squared forecast errors, [31].

3.3. AR and ARIMA processes

Box et al. in [11] propose a general framework based on an autoregressive integrated moving average (ARIMA) process of order (p,d,q) to model stationary and nonstationary time series. The process can be expressed by:

$$\phi(B)(1-B)^d y_t = \theta(B)a_t, \tag{4}$$

where y_t is an observable time series and a_t is a white noise process having mean zero and variance σ_a^2. The backward shift operator is denoted by $Bz_t = z_{t-1}$. The Autoregressive and Moving Average operators are defined by $\phi(B)$ and $\theta(B)$ polynomials of order p and q respectively and d denotes the order of differencing that is required to make the time series stationary.

The automatic identification procedure consists of selecting the best ARIMA model from a full range of possibilities according to the Akaike Information Criterion (AIC), [32]. The AIC combines the fit of the model with the number of parameters used in order to avoid over parameterisation, and can be expressed normalized by sample size n as follows:

$$AIC_{p,q} \approx ln(\hat{\sigma}_a^2) + r\frac{2}{n}, \tag{5}$$

where $r = p + q$. The models estimated include orders: i) p=1,2,3; ii) q=1,2,3; and iii) d=1,2, since higher p, q, d values are not generally used in practice, [11]. The preferred model is the one with the minimun AIC value.

A simpler form of the model involves only the identification of the autoregressive part, which is essentially a dynamic regression on past lags of the time series. The identification of the AR order is done again by AIC optimisation and the model assumes stationarity of the time series.

3.4. ARX models

An ARX model structure can be expressed by a linear difference equation:

$$y_t + a_1 y_{t-1} + \ldots + a_{n_a} y_{t-n_a} = b_1 u_{t-n_k} + \ldots + b_{n_b} u_{t-n_k-n_b+1} + a_t, \tag{6}$$

where AR refers to the autoregressive part and X to the extra input, sometimes called the exogenous variable. The parameters n_a and n_b are the orders of the ARX model, and n_k is the number of input samples that occur before the input affects the output, also called the dead time in the system [33]. The variables y_t and u_t stand for the volume received and retailer sales, respectively. Model orders n_a, n_b and n_k have been chosen by minimizing the AIC. Model selection and the estimation of the unknown parameters a_i, $i = 1, \ldots, n_a$ and b_j, $j = 1, \ldots, n_b$ have been done by means of the routines implemented in the System Identification toolbox ($MATLAB^{TM}$)

3.5. Neural Networks

Artificial Neural Networks (NN) have been successfully applied in both univariate and multivariate time series forecasting [34]. It has been long proposed that their use can be beneficial to supply chain modelling and specifically for forecasting [35]. In [36] NN were shown to be more accurate than MA and regression, however with no statistically significant differences, when applied to supply chain forecasting problems. The motivation for using NN in this study arises from their assumption-free modelling. They do not require imposing any assumptions regarding the data generating process of the time series or interactions, if any, between variables. The mapping between inputs and outputs is data-driven, limiting the modelling decisions required by the human expert [34], thus allowing us to identify empirically any existing nonlinearities in the dataset.

The most widely employed architecture is the multilayer perceptron (MLP), which we will be employing in this study. These are well researched regarding their properties and have been shown to be able to generalise any linear or nonlinear functional relationship between the inputs and the outputs, to any degree of accuracy without any prior assumptions about the underlying data generating process [34, 37]. In principle these can be seen as nonlinear extensions of AR and ARX models.

MLPs are feed-forward networks that are organised in layers of nodes, that control the information flow within the model. The first layer is the input layer. This is followed by any number of hidden layers, where the main part of the processing is performed. A hidden layer is made of any number of hidden nodes. A hidden node can be linear or nonlinear. If it is linear then each node is identical to a multiple linear regression, however typically they are nonlinear, but able to approximate both linear and nonlinear behaviours. It has been argued that a single hidden layer is adequate for most forecasting purposes and the modeller has to identify the correct number of hidden nodes [34]. The last layer of the network, the output layer, adds the information from the preceding layers to the required output. Typically this is done through a linear combination of the outputs of the nodes of the previous layers. An example of the architecture of a MLP with three inputs, a single hidden layer with four nodes and a single output can be seen in Figure 6.

Given a time series y, at a point in time t, a one-step ahead forecast F_{t+1} is computed using $p = I$ inputs that can be lagged observations of y_t or explanatory variables, lagged or not. I denotes the number of input units p_i of the NN. The functional forms is

$$F_{t+1} = \beta_0 + \sum_{h=1}^{H} \beta_h g\left(\gamma_{0i} + \sum_{i=1}^{I} \gamma_{hi} p_i\right). \tag{7}$$

where $w = (\beta, \gamma)$ are the network weights and $\beta = [\beta_1, \ldots, \beta_H]$, $\gamma = [\gamma_{11}, \ldots, \gamma_{HI}]$ are for the output and the hidden layers respectively. The β_0 and γ_{0i} are the biases of each neuron, which are identical in function to the constant in linear regression. I and H are the number of input and hidden nodes in the network and $g(\cdot)$ is a nonlinear transfer function, which is usually either the sigmoid logistic or the hyperbolic tangent function [38] and provides the nonlinear capabilities to the model. MLPs offer extensive degrees of freedom in modeling for prediction tasks. The modeler must choose the appropriate data and its pre-processing, the NN architecture, the signal processing within nodes, the training algorithm and the cost function. For a detailed discussion of these issues and the ability of NNs to forecast time series, the reader is referred to [34].

In this analysis we develop both univariate and multivariate networks. The networks use the inputs identified for the AR and ARX models discussed before. The rest of the parameters of the networks are identified through simulation and are provided in Table 1. The univariate model is named NAR and the multivariate is named NARX. Furthermore, we provide the results for model NARX-Lin which involves direct connections of the inputs to the output layer, as well as through the hidden node, thus achieving direct modelling of both linear and nonlinear information. The model is formulated as:

$$F_{t+1} = \beta_0 + \sum_{h=1}^{H} \left(\beta_h g\left(\gamma_{0i} + \sum_{i=1}^{I} \gamma_{hi} p_i\right)\right) + \sum_{i=1}^{I} \delta_i p_i, \tag{8}$$

where δ_i are the connecting weights between the inputs and the output node, which is linear. Results for a univariate NAR-Lin model are not provided since there was no accuracy gains over the NAR model. Note that NARX is in theory able to approximate NARX-Lin [37], however in practice formulating explicitly the shortcut connections between the input and the output layer, as in NARX-Lin, aids significantly the training of the model.

All networks use for their training Bayesian Regularisation and no validation set is needed as in typical NN training [39]; therefore we use exactly the same data for training and evaluating the NNs as for the other models. All models use the sigmoid logistic function for their hidden layers:

$$f(p) = 1/(1 + exp^{-p}), \tag{9}$$

where p is an input. The networks are built using the Neural Network toolbox in $MATLAB^{TM}$ using standard functions.

4. Empirical results

In this section predictive validation is used to compare models. For this purpose, 20% of the data constituted by the last 10 weeks of each SKU are kept for comparing the performance of the proposed methods, as hold-out sample. These last 10 weeks are not used for the parameter estimation of the models, but retained to produce rolling one-step ahead out-of-sample forecasts, without refitting the models, which are used to evaluate the performance of the alternative forecasting methods. The results are first analysed by forecasting accuracy, assessing whether the methods that include downstream information are more accurate or not. Afterwards, encompassing tests are carried out to identify if the multivariate methods add significantly more information in comparison to the univariate methods.

4.1. Out-of-sample forecasting performance

Forecast errors are measured across time for each SKU by means of the Mean Absolute Percentage Error (MAPE) and the Median Absolute Percentage Error (MdAPE), i.e:

$$\begin{aligned} MAPE &= mean(|PE_t|), \\ MdAPE &= median(|PE_t|), \end{aligned}$$

where PE_t is the percentage error given by $PE_t = 100(y_t - F_t)/y_t$, $t = 1, \ldots, N$. Here, y_t stands for the actual value and F_t is the forecast, both of them at time t. The MdAPE is a more robust implementation of MAPE in presence of outliers, [40].

These measures are chosen due to their simplicity of interpretation and applicability to this particular dataset. A rolling origin one-step ahead forecast is produce for each of the 10 weeks in the out of sample. The percentage errors of these forecasts are used to calculate the MAPE and MdAPE of each individual SKU across the different time origins, which are afterwards aggregated in dataset average figures, obtaining the Mean(MAPE), Mean(MdAPE) as overall error measures over all SKUs. These latter measures will be used to compare forecasting accuracy between the forecasting methods.

Table 2 shows the Mean of the MAPE and MdAPE of the considered methods for the out-of-sample evaluation. In this table the minimum values per measure are highlighted in bold. We can easily observe that the multivariate methods are more accurate than the univariate ones. This indicates that information sharing reduces the forecast errors on average. Note that AR, NAR, MA, SES, and ARIMA obtain similar error level. Regarding the forecast error variability measured by the standard deviation provided (St. Dev.) in Table 2, it is apparent that the multivariate models on average exhibit lower dispersion, with the lowest achieved by ARX. Across the univariate models the lowest error variability is achieved by the nonlinear NAR. The same conclusions can be obtained from the forecast error boxplots of the MAPE and MdAPE across SKUs depicted in Figure 7. We provide the mean error on the same figure as well. Again, the multivariate models show better performance in comparison to the rest of the techniques and less dispersion. The percentage of SKUs that collaboration leads to improvement is 65.1% and 62.8% for MAPE and MdAPE respectively.

We have performed statistical tests to identify whether the reported accuracy differences are significant. To avoid any assumptions of normality we employ a series of non-parametric tests. Initially, we use the one-way Friedman tests, which is the non-parametric analogous to the ANOVA test; testing if at least one of the methods is performing significantly different from the rest. For all MAPE, MdAPE and St. Dev there are significant differences with reported p-values equal to 0. To clarify which methods perform significantly different we use the Nemenyi post-hoc test. This is again a non-parametric test, based on calculating the mean rank of each method. A critical distance for the set of methods compared is computed and any methods within this critical distance have no significant differences. The reader is pointed for more information to [41]. The numerical results of the non-parametric tests are provided in Table 3, whereas a visualisation of the results of the Nemenyi test is provided in Figure 8.

The statistical tests indicate clearly that that the results can be separated into two groups of methods; the univariate and the multivariate. There are no statistically significant differences in accuracy for both MAPE and MdAPE across the multivariate methods; hence it is impossible to conclude that one of these methods performs better. Among the univariate methods there is a similar picture, with the exception of the Naïve method that significantly underperforms compared to NAR, SES and ARIMA.

Therefore, from these results we can claim that sharing information reduces the forecast error mean and variability and is beneficial.

4.2. Encompassing tests

A forecast encompassing test allows us to evaluate whether a forecasting method contains more valuable forecasting information compared to another method, or simply if a method encompasses another. This way we can test the hypothesis if the univariate models are encompassed by the multivariate models that make use of the information sharing, i.e. they contain more valuable information, or not; hence providing further evidence of the benefits of such a process. There are a number of models that can be used as the basis of encompassing tests [42]. The test we use is based on:

$$y_t = \alpha_0 + \alpha_1 F_{1t} + \alpha_2 F_{2t} + e_t, \qquad (10)$$

where F_{1t} and F_{2t} are the forecasts of two methods, α_0 is a constant that permits the possibility of bias and y_t is the observation at time t. Equation (10) can be examined either in an unconstrained or a constrained form, where in the latter $\alpha_1 + \alpha_2 = 1$. Here we use the latter, since without the constraint the results show little more than the possible collinearity of the methods [43].

Table 4 presents the results of the encompassing tests. We provide the p-value of each combination of models. Combinations of models with insignificant contributions are in boldface. In this table we want to evaluate whether the multivariate models offer additional useful information to the univariate models, indicating a beneficial effect of information sharing and also to examine whether the univariate models capture additional information, in comparison to the multivariate models. From Table 4 we can conclude that all ARX, NARX and NARX-Lin contribute significantly to the univariate models, providing further evidence of the importance of information sharing. Considering the multivariate methods, only NAR, MA and the Naïve methods contribute to ARX, but not to NARX or NARX-Lin. In closer examination we can see that the p-values of the univariate models are close to 0.05, implying weak evidence of contribution to ARX, which we would reject under 1% significance level. Moreover, this is interpretted as evidence of possible nonlinearities in the supplier time series that are captured by the nonlinear part of NAR, NARX and NARX-Lin.

Furthermore, we can observe that the inclusion of the direct linear modelling of information with the NARX-Lin method allows capturing both linear and nonlinear components explicitly, thus contributing to ARX, providing further evidence of nonlinearities. From these results we conclude that the univariate methods have not captured additional significant information over the multivariate, whereas the opposite is true. The multivariate encompass the univariate ones and provide significant improvements due to their access to downstream information of the supply chain.

5. Conclusions

The utility of information sharing with regards to forecasting performance is a controversial issue. Theoretical analysis relying on restrictive assumptions claims that the information available in the market sales can be extracted by the upstream level in the supply chain by filtering the retailer orders signal. Therefore, market sales information sharing would not bring significant improvements in terms of forecasting accuracy. On the other hand, empirical analysis accomplished in particular companies reached different conclusions. Mainly, they see a clear benefit of sharing information. Nonetheless, the number of case studies is still small.

The results of this paper conclude that information sharing improves forecasting performance, resulting in 6 to 8 percentage points lower forecasting error as measured by MdAPE and MAPE respectively. That result was based on the benchmarking of multivariate against univariate models using a real dataset, based on a serially linked supply chain. Automatic system identification techniques were employed to model the supplier demand. In addition, no restrictions about either promotions, replenishment rules or demand were imposed. Statistical tests indicated significant gains in forecasting accuracy of the multivariate models over the univariate models, demonstrating a clear benefit of information sharing for reducing forecasting errors. Furthermore, we employed forecast encompassing tests to identify whether there is significant information that was missing in either uni- or multivariate models and concluded that the multivariate models contributed significantly to all univariate models, while the opposite was not true. This provides further empirical evidence of the importance of information sharing.

Crucially, this study has added to the research on information sharing by demonstrating for a range of SKUs that multivariate methods can lead to substantial improvements in forecasting accuracy. This gives manufacturers the incentive to explore the improvements they can achieve within their own supply chain. The gains will of course depend on the nature of the supply chain and in particular the process and information through which the retailer collaborates. This raises a wealth of interesting questions about the type of information that is most valuable. Furthermore, there are clear managerial implications. A manufacturer has direct interest in information sharing, as it can bring significant monetary savings, enhanced customer service levels and lower inventory costs. Retailers downstream have indirect benefits from this collaboration. As the manufacturer enjoys better inventory management, the retailers can benefit from higher service level and potentially reduced buying costs, thus providing further incentives for managers at every level of the supply chain to collaborate.

While the forecast improvements are substantial compared with alternatives as enhanced statistical forecasting based only on internal company data, they come at a cost. Implementation issues include enhanced software and better trained staff as well as a more extensive forecasting process that includes the collection and interpretation of the valuable downstream information.

This study has focused on providing solid empirical evidence whether information sharing affects the upstream forecasting accuracy. Although our findings imply that information sharing mitigates the Bullwhip Effect (we provide indirect evidence through the reduction of forecasting errors), this work does not provide a causal connection between Bullwhip, its intensity and the benefits of information sharing and therefore falls short of showing whether the increase in forecasting accuracy is due to reduction of the Bullwhip Effect, or part of it, or other unexplored reasons. In other words, whereas there is a general agreement regarding the harmful consequences of the Bullwhip Effect to forecasting accuracy in upstream levels, the quantification and magnitude of that connection is not well defined.

Future studies should quantify the cost reduction achieved by improving forecast accuracy to give more solid reasons to practitioners whether to invest on inter-organizational information systems and advanced Forecasting Support Systems. Moreover, further research can be addressed on extending the case study to different industries, particularly for those that experience different degrees of Bullwhip Effect, providing additional evidence of the connection between forecasting accuracy benefits and information sharing.

Acknowledgment

The authors thank the referees for valuable comments on an earlier version of this paper. J. R. Trapero is supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme.

References

[1] S. Geary, S. M. Disney, D. R. Towill, On bullwhip in supply chains-historical review, present practice and expected future impact, International Journal of Production Economics 101 (2006) 2–18.

[2] H. L. Lee, V. Padmanabhan, S. Whang, The bullwhip effect in supply chains, Sloan Management Review Spring (1997) 93–102.

[3] H. L. Lee, V. Padmanabhan, S. Whang, Information distortion in a supply chain: the bullwhip effect, Management Science 43 (1997) 546–558.

[4] S. Cannella, E. Ciancimino, On the bullwhip avoidance phase: supply chain collaboration and order smoothing, International Journal of Production Research 1 (2009) 1–38.

[5] D. Seifert, Collaborative planning, forecasting, and replenishment: how to create a supply chain advantage, New York: AMACOM, 2003.

[6] P. Fiala, Information sharing in supply chains, Omega 33 (5) (2005) 419–423.

[7] M. Holweg, S. M. Disney, J. Holmström, J. Småros, Supply chain collaboration: making sense of the strategy continuum, European Management Journal 23 (2005) 170–181.

[8] C. Zhang, G.-W. Tan, D. J. Robb, X. Zheng, Sharing shipment quantity information in the supply chain, Omega 34 (2006) 427–438.

[9] S. Raghunathan, Information sharing in a supply chain: A note on its value when demand is nonstationary, Management Science 47 (2001) 605–610.

[10] D. Towill, A. D. Vecchio, The application of filter theory to the study of supply chain dynamics, Production Planning and Control 5 (1994) 82–96.

[11] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, Time series analysis: Forecasting and Control. (3rd. Edition), Upper Saddle River, New Jersey: Prentice Hall, 1994.

[12] T. Hosoda, M. M. Naim, S. M. Disney, A. Potter, Is there a benefit to sharing market sales information? Linking theory and practice, Computers & Industrial Engineering 54 (2008) 315–326.

[13] T. Hosoda, S. M. Disney, On variance amplification in a three-echelon supply chain with minimum mean square error forecasting, Omega 34 (4) (2006) 344 – 358.

[14] K. K. Kim, S. Y. Ryoo, M. D. Jung, Inter-organizational information systems visibility in buyer-supplier relationships: The case of telecommunication equipment component manufacturing industry, Omega 39 (2011) 667–676.

[15] P. J. Byrne, C. Heavey, The impact of information sharing and forecasting in capacited industrial supply chains: a case study, International Journal of Production Economics 103 (2006) 420–437.

[16] R. Fildes, P. Goodwin, M. Lawrence, K. Nikolopoulos, Effective forecasting and jugdmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning, International Journal of Forecasting 25 (2009) 3–23.

[17] P. Goodwin, Integrating management judgment and statistical methods to improve short-term forecasts, Omega 30 (2002) 127–135.

[18] R. Fildes, P. Goodwin, Against your better judgment? how organizations can improve their use of management judgment in forecasting, Interfaces 37 (2007) 70–576.

[19] P. H. Franses, R. Legerstee, Do experts adjustments on model-based SKU-level forecasts improve forecast quality?, Journal of Forecasting 29 (2010) 331–340.

[20] T. Kelepouris, P. Miliotis, K. Pramatari, The impact of replenishment parameters and information sharing on the bullwhip effect: a computational study, Computers and Operation Research 35 (2008) 3657–3670.

[21] L. Lapide, New developments in business forecasting, The Journal of Business forecasting 20 (4) (2001) 11, 12, and 36.

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

[23] J. Dejonckheere, S. M. Disney, M. R. Lambrecht, D. R. Towill, Measuring and avoiding the bullwhip effect: A control theoretic approach, European Journal of Operational Research 147 (2003) 567–590.

[24] V. Shukla, M. M. Naim, E. A. Yaseen, "Bullwhip" and "backlash" in supply pipelines, International Journal of Production Research 47 (2009) 6477–6497.

[25] J. C. Fransoo, M. J. F. Wouters, Measuring the bullwhip effect in the supply chain, Supply Chain Management: An International Journal 5 (2) (2000) 78–89.

[26] R. J. Hyndman, A. B. Koehler, R. D. Snyder, S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, Int. J. Forecast. 18 (3) (2002) 439–454.

[27] S. Makridakis, S. C. Wheelwright, R. J. Hyndman, Forecasting: Methods and applications. Third Edition, John Wiley & Sons, 1998.

[28] R. J. Hyndman, A. B. Koehler, J. K. Ord, R. D. Snyder, Forecasting with Exponential Smoothing: The State Space Approach, Berlin, Heidelberg: Springer-Verlag, 2008.

[29] E. S. Gardner, Exponential smoothing: The state of the art, Part II, International Journal of Forecasting 22 (2006) 637–666.

[30] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, R. Parzen, R. Winkler, The accuracy of extrapolation (time series) methods; results of a forecasting competition, Journal of Forecasting 1 (1982) 111–153.

[31] E. S. Gardner, Exponential smooothing: The state of the art, Journal of Forecasting 4 (1) (1985) 1–28.

[32] H. Akaike, A new look at the statistical model identification, IEEE Transaction on Automatic Control 19 (6) (1974) 716–723.

[33] L. Ljung, System Identification, Theory for the user, 2nd Edition, Upper Saddle River, NJ: PTR Prentice Hall, 1999.

[34] G. Q. Zhang, B. E. Patuwo, M. Y. Hu, Forecasting with artificial neural networks: The state of the art, International Journal of Forecasting 14 (1) (1998) 35–62.

[35] H. Leung, Neural networks in supply chain management, in: Engineering Management Conference. 'Global Engineering Management: Emerging Trends in the Asia Pacific', Proceedings of IEEE Annual International, 1995, pp. 347 –352.

[36] R. Carbonneau, K. Laframboise, R. Vahidov, Application of machine learning techniques for supply chain demand forecasting, European Journal of Operational Research 184 (3) (2008) 1140–1154.

[37] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Networks 4 (2) (1991) 251–257.

[38] U. Anders, O. Korn, Model selection in neural networks, Neural Networks 12 (2) (1999) 309–323.

[39] D. MacKay, Bayesian interpolation, Neural Computation 4 (3) (1992) 415–447.

[40] R. Fildes, The evaluation of extrapolative forecasting methods, International Journal of Forecasting 8 (1992) 81–98.

[41] J. Demsar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.

[42] Y. Fang, Forecasting combination and encompassing tests, International Journal of Forecasting 19 (1) (2003) 87 – 94.

[43] Y. Y. Chong, D. F. Hendry, Econometric evaluation of linear macro-economic models, The Review of Economic Studies 53 (4) (1986) 671–690.

FIGURES

List of Figures

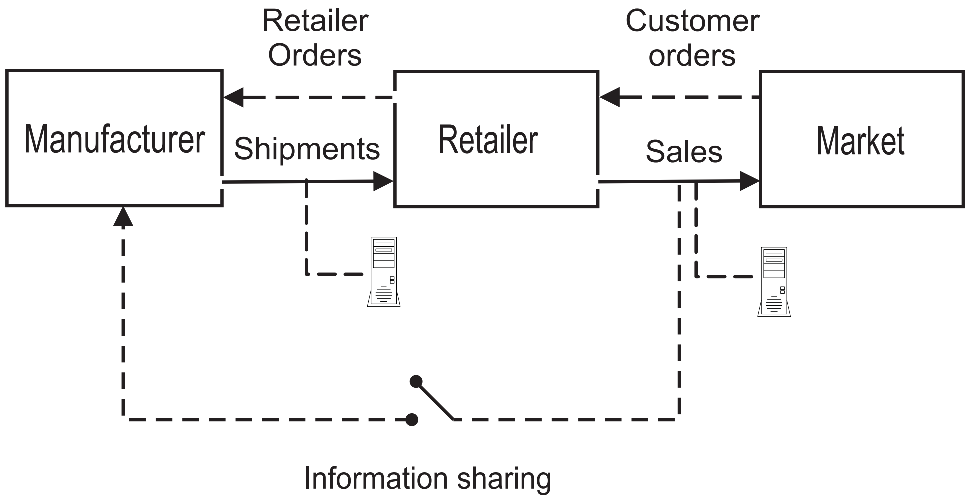


Figure 1: Flow of material (—) and information (- -) for a two serially linked supply chain.

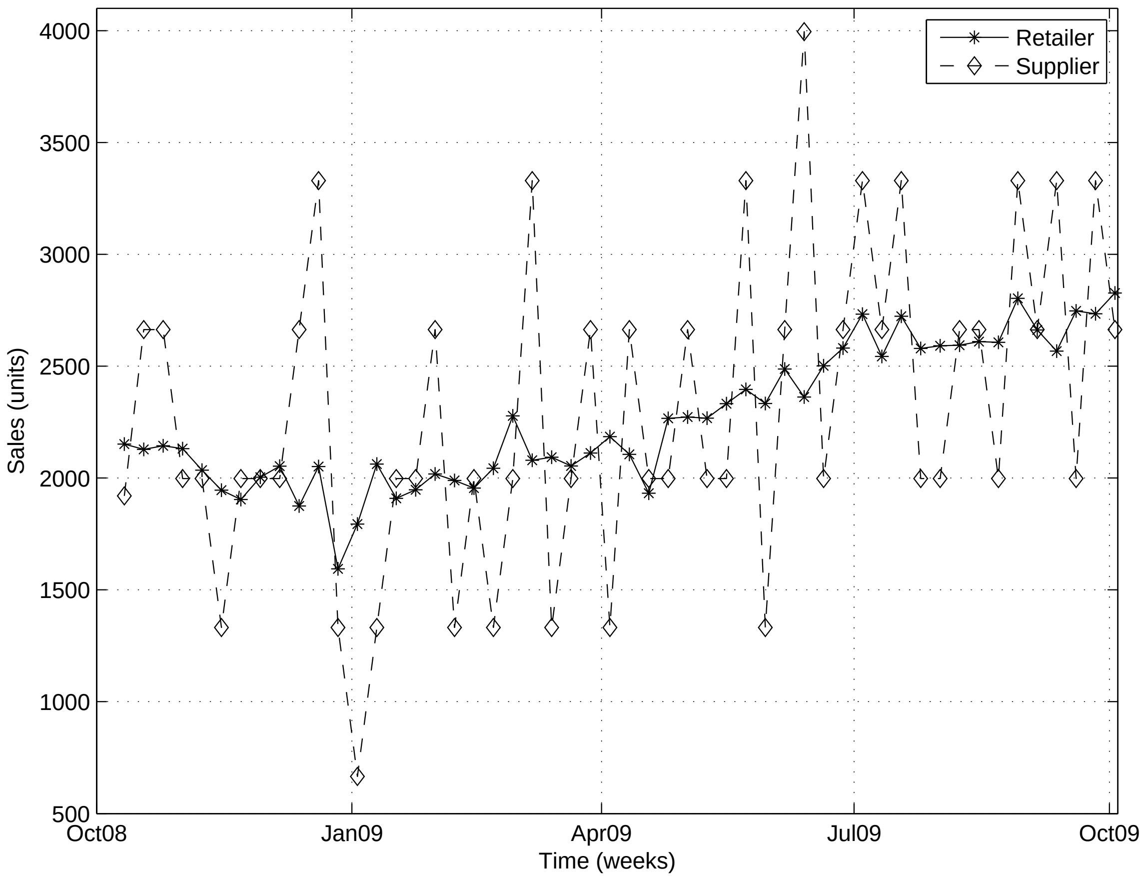


Figure 2: Example of a typical SKU. Retailer sales are in a solid line (—) and Volume received in a dashed line (- -)

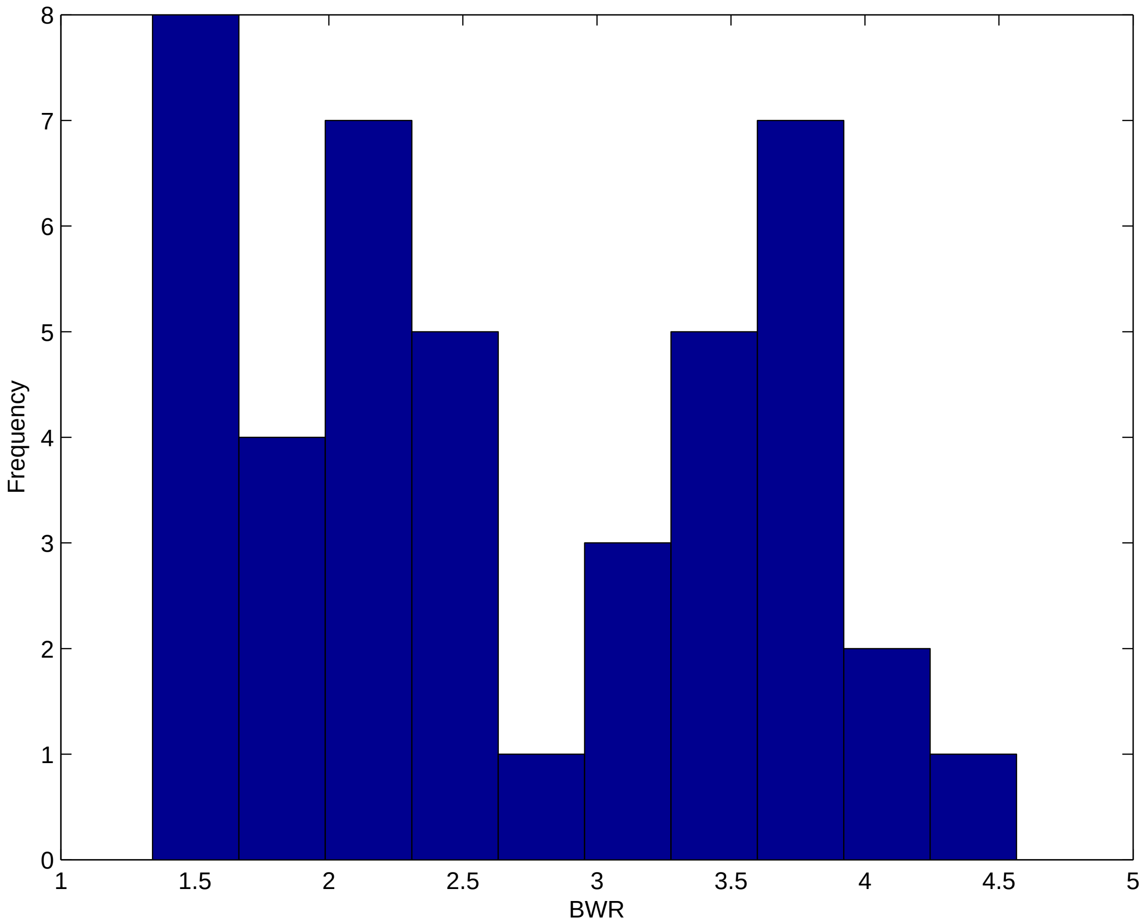


Figure 3: Histogram of bullwhip ratio (BWR).

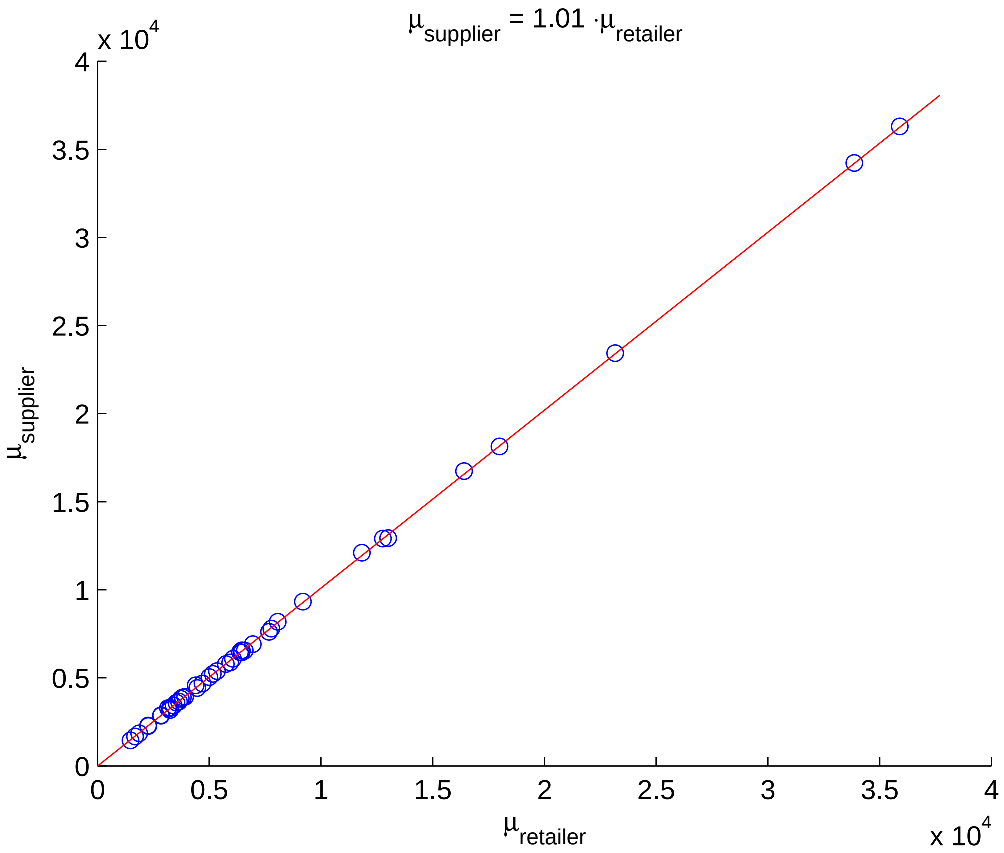


Figure 4: Scatter plot between retailer and supplier mean sales.

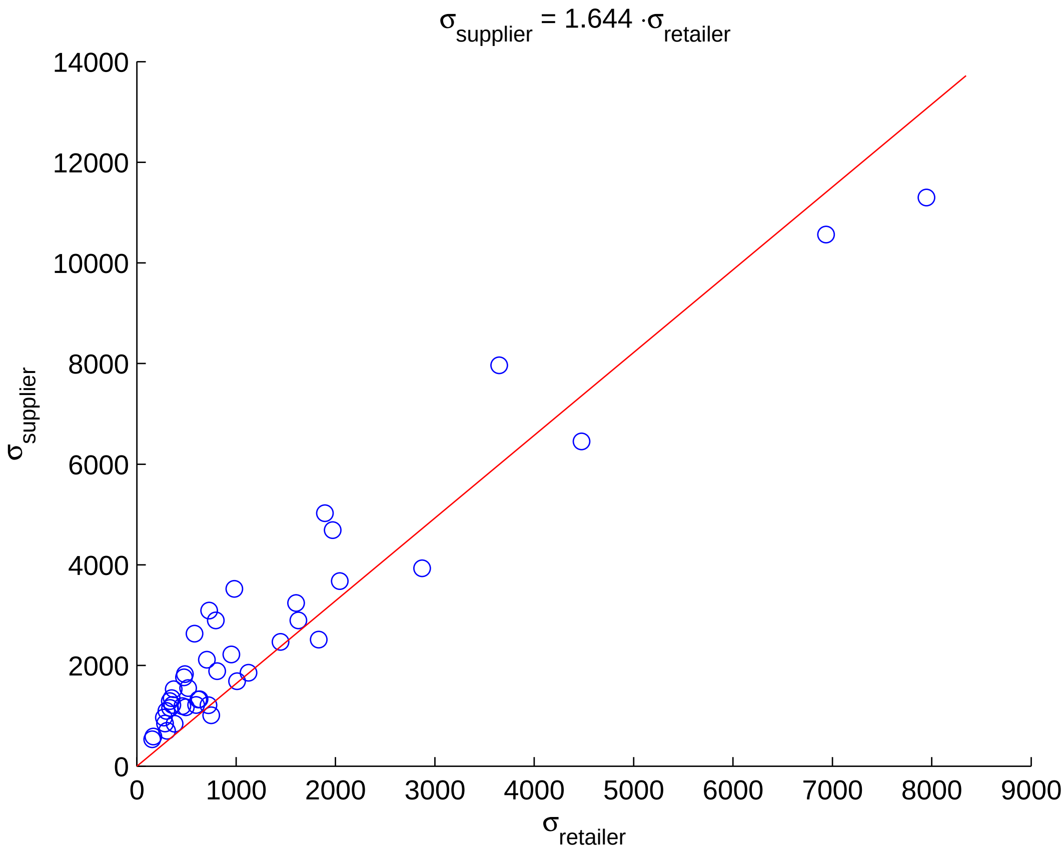


Figure 5: Scatter plot between retailer and supplier sales standard deviation.

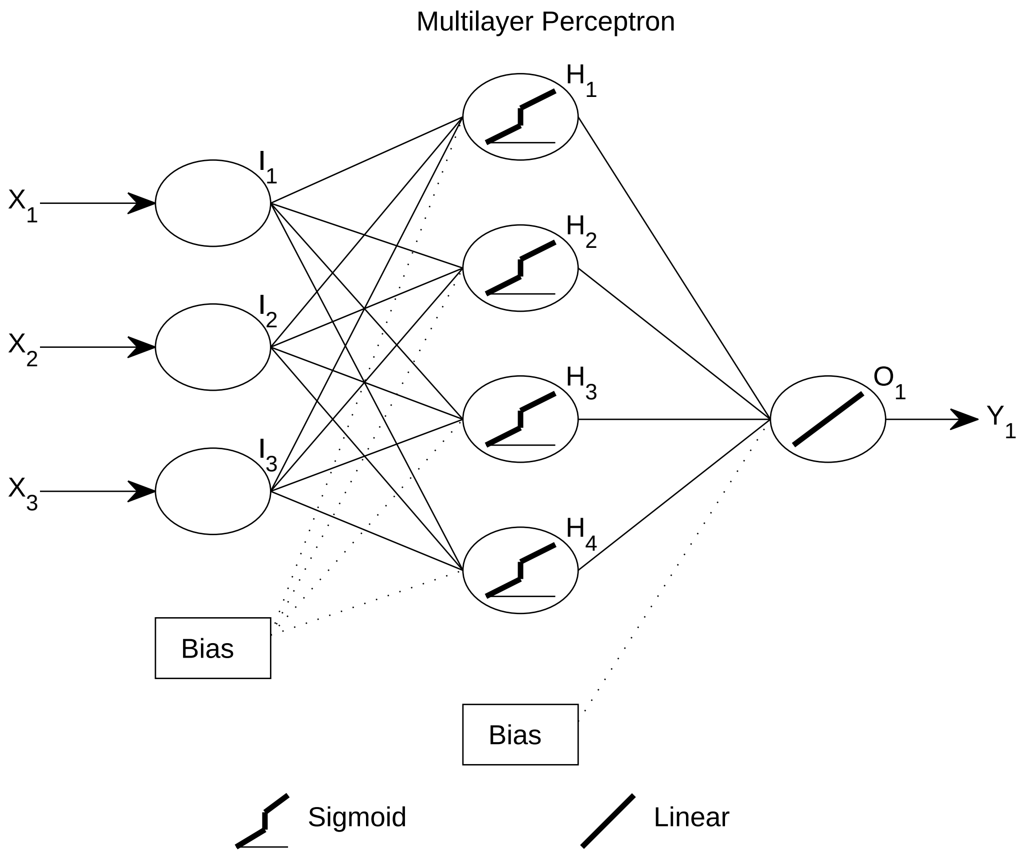
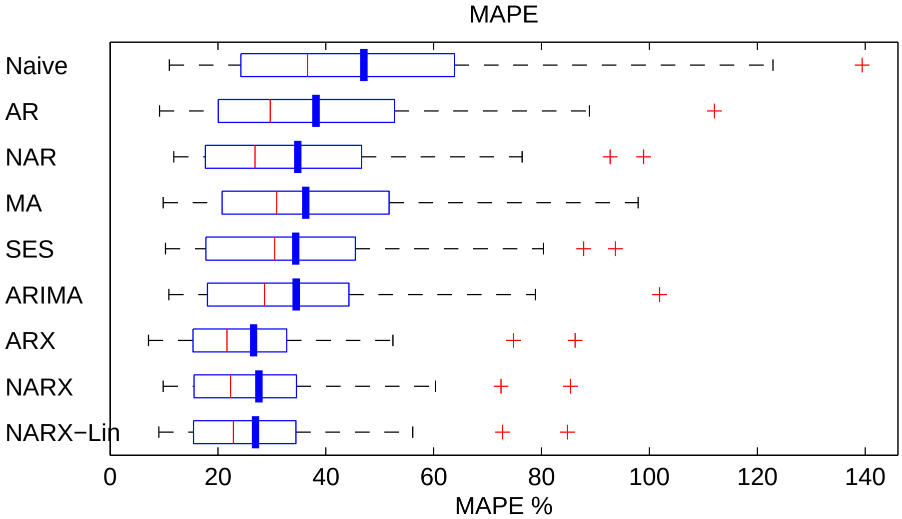


Figure 6: An example MLP setup with 3 inputs and 4 hidden nodes.



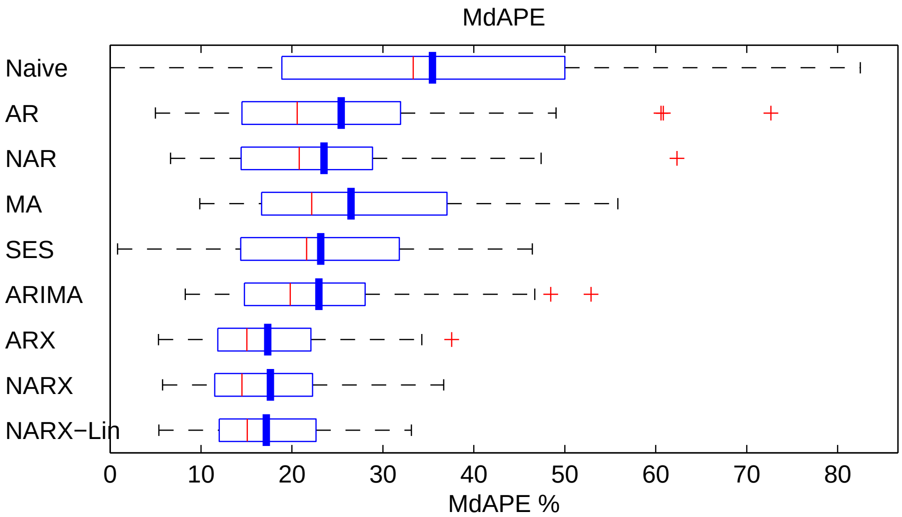


Figure 7: MAPE and MdAPE boxplots for forecasting methods. The thick blue line represents the mean of each error distribution.

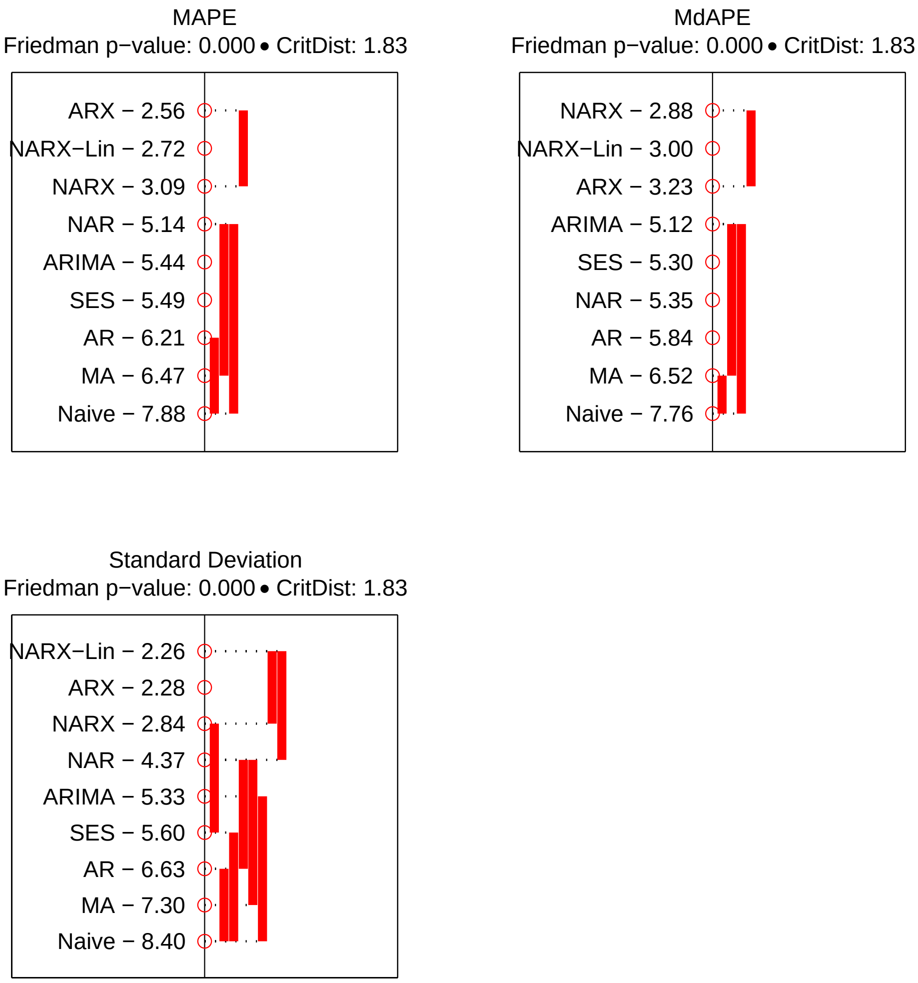


Figure 8: Methods are sorted from best (top) to worst (bottom). For methods joined by a red line there is no significant evidence of accuracy difference at 5% level. Multiple lines correspond to using different forecasting methods for evaluating the Nemenyi's test critical distance centre.

TABLES

List of Tables

Table 1: Neural Network models design parameters.

Model Name	Hidden Nodes	Bias (Hidden, Output)	Training Epochs	Scaling
NAR	1	No, Yes	2000	[-0.75, 0.75]
NARX	8	Yes, No	2000	[-0.75, 0.75]
NARX-Lin	11, 1	No, No, No	2000	[-0.75, 0.75]

Table 2: Mean of the MAPE %, MdAPE % and standard deviation of the residuals for all forecasting methods.

Method	Univariate						Multivariate		
	Naïve	AR	NAR	MA	SES	ARIMA	ARX	NARX	NARX-Lin
MAPE %	47.05	38.20	34.82	36.28	34.43	34.50	**26.63**	27.61	26.97
MdAPE %	35.46	25.41	23.53	26.50	23.16	22.96	17.35	17.62	**17.18**
St. Dev.	3183.67	2562.06	2175.20	2555.98	2285.67	2269.31	**1880.92**	1922.75	1888.70

Lowest figure in each row is in boldface.

Table 3: Friedman and Nemenyi tests results.

Metric	Friedman p-value	Nemenyi Mean Rank*								
		Naïve	AR	NAR	MA	SES	ARIMA	ARX	NARX	NARX-Lin
MAPE %	0.000	7.88	6.21	5.14	6.47	5.49	5.44	2.56	3.09	2.72
MdAPE %	0.000	7.76	5.84	5.35	6.52	5.30	5.12	3.23	2.88	3.00
St. Dev.	0.000	8.40	6.63	4.37	7.30	5.60	5.33	2.28	2.84	2.29

*Lowest mean rank is better; Critical distance for Nemenyi test are 2.12, 1.83 and 1.69 for 1%, 5% and 10% significance level respectively.

Table 4: Encompassing tests results.

p-value		Method II								
		Naïve	AR	NAR	MA	SES	ARIMA	ARX	NARX	NARX-Lin
Method I	Naïve	-	0.001	0.004	**0.155**	**0.314**	0.027	0.043	**0.267**	**0.116**
	AR	0.000	-	0.000	0.000	**0.071**	**0.112**	**0.864**	**0.610**	**0.776**
	NAR	0.000	0.000	-	0.000	0.005	0.000	0.045	**0.490**	**0.190**
	MA	0.000	0.000	0.000	-	**0.545**	0.000	0.021	**0.145**	**0.062**
	SES	0.000	0.000	0.000	0.000	-	0.000	**0.643**	**0.704**	**0.374**
	ARIMA	0.000	0.000	0.000	0.000	0.000	-	**0.729**	**0.721**	**0.930**
	ARX	0.000	0.000	0.000	0.000	0.000	0.000	-	0.000	0.002
	NARX	0.000	0.000	0.000	0.000	0.000	0.000	**0.269**	-	**0.483**
	NARX-Lin	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.000	-

The p-value shows whether method I contributes significantly to method II. Insignificant contributions at 5% level are shown in boldface.