Lancaster University Management School
Working Paper
2010/044

Impact of Information Exchange on Supplier Forecasting Performance

Juan Trapero, Nikolaos Kourentzes and Robert Fildes

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

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. This works assesses the role of sharing market sales information obtained by the retailer on the supplier forecasting accuracy. Data have been collected from a manufacturer of domestic cleaning products and a major UK grocery retailer to show the circumstances where information sharing leads to improved accuracy. We find significant evidence of benefits through information sharing.

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)

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 in [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. In fact, information transparency is one of the ten principles proposed in [1] to achieve bullwhip reduction.

Holweg et al. in [6] suggest a classification of four different supply chain collaborations depending on the extent in planning collaboration 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. However, there is no general agreement from the literature at this point. In fact, some authors, based on analytical models claim that all the information available in the market sales can be translated upwards the stages of a supply chain by the subsequent orders. Thus, a chain echelon might retrieve such information by means of a filter. However, in order to make the problem mathematically tractable those works rely on restrictive assumptions that tend to be highly modified versions of reality [7]. For instance, they do not include the influence of promotions. However, according to the companies studied in [8] and [9] promotions appear at 90 and 60 percent of observations, respectively. Therefore, insights obtained from theoretical developments are limited [7].

On the other hand, some authors based on real customer demand datasets state that information sharing improve the forecasting accuracy [7], [10], [11]. In fact, an attempt to minimize the gap between theory and practice was done in [11] by analyzing the influence of supply chain collaboration on the basis of real data. Basically, an analysis of three SKUs with different level of sales and without promotion sales were considered. Furthermore, they limit the real customer demand and orders to models that follow an AR(1) and ARMA(1,1) structure, respectively raising the question as to whether this is too limiting a specification to understand the benefits achievable.

In this article, we extend the number of SKUs analyzed in our test dataset, as well as, we neither do not assume any restriction about promotions nor any particular structure for the customer demand. In order to achieve more generic conclusions we use automatic system identification procedures to select the adequate structure for the supplier sales using the retailer sales information. We employ both linear and nonlinear AR models with explanatory (exogeneous) variables (ARX models). Looking forward, the results show that the supplier can improve the forecast performance by using the market sales shared by the retailer. Other univariate techniques such as ARIMA, exponential smoothing, Moving Averages and Neural Networks where employed as benchmarks. The supply chain characteristics of the bullwhip offer only a partial explanation of the relative errors.

This 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, while section 5 provides a discussion of the results. Finally, main conclusions are drawn in Section 6.

2. Case study

The supply chain system consists of a serially linked two-level supply chain, see Figure 1. There is a flow of information from the market towards the manufacturer and an inverse one regarding materials. 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. When it is on, market sales information is available for the manufacturer.

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 UK. 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 [11] use the retailer orders as a measure of supplier sales. Given that we just had available information about volume received by the retailer, we employed that volume as a measure of supplier sales. The volume received by the retailer is a delayed version of the retailer orders time series.

In summary, 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 standard deviations between the output supplier sales and the input retailer sales. Let the Bullwhip Ratio (BWR) be denoted by:

$$BWR = \sigma_{supplier}/\sigma_{retailer}. \quad (1)$$

Dejonckheere et al. in [12] propose other two bullwhip measures based on the frequency response plot (FR). However, in order to compute the frequency response plot 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). However, it should be pointed out that if extreme (very high or low) sales are expected as a consequence of a promotion campaign, (1) may be estimated by means of the Median Absolute Deviation. This latter statistical measure provides a more robust version less sensitive to extreme observations. The above measure has a significant weakness for time series data. If either time series of the supplier or the retailer sales exhibit trend (local or global) or seasonality the estimation of $\sigma_{supplier}$ or $\sigma_{retailer}$ does not measure only the variability around the level of the time series that one would need, but additional variability due to the nonstationarities in the time series.

To overcome these limitations we propose a new method to measure the Bullwhip Ratio. Given a time series y, we fit a Least Absolute Deviation (LAD) regression and calculate the Root Mean Squared Error of the residuals, which essentially measures the deviation around a robust estimation of the level of the time series, instead of the mean $\overline{y}$ as in the normal standard deviation calculation. The LAD regression is similar to normal OLS regression, with the exception that the cost function is based on absolute instead of squared errors, therefore it is more robust to outlying values, such as promotions. We denoted this as $\sigma_{y_{LAD}}$. This new estimator is robust to outliers and additionally can follow changes in the level of the time series, thus resulting in more informative measurement of BWR. We propose:

$$Robust\ BWR = \sigma_{supplier_{LAD}} / \sigma_{retailer_{LAD}}, \tag{2}$$

as a robust estimation of the Bullwhip Ratio, which we will use in the analysis that follows.

Figure 3 shows the histogram of BWR calculated using (1) and (2). In this figure we can see that the two measures provide different results, with the Robust BWR showing higher measured ratios, due to the more robust estimation of the time series deviation around the mean. The differences between the two measures are primarily due to the differences between $\sigma_{retailer}$ and $\sigma_{retailerLAD}$. Furthermore we can observe that under both estimations 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, depending on the calculation. Note that some SKUs can reach a BWR greater than 4.

According to Figure 3 we can classify the SKUs in our dataset depending on the BWR. Considering the Robust BWR and taking a value of 3.24 as a threshold, SKUs with BWR bigger than 3.24 are denoted as High BWR group and those SKUs lower than 3.24 correspond to Low BWR. Table 1 shows some descriptive statistics. Observe that for both measurements the number of High BWR SKUs is lower than the Low BWR ones.

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. However, what if we analyze 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 verifies 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. On the other hand, 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.

3.1. Naïve and Moving Average

In order to reduce the time series random variation and to extract the low-frequency trend-cycle component a moving average can be used. A moving average forecast of order k, or $MA(k)$, is given by:

$$F_{t+1} = \frac{1}{k} \sum_{t-k+1}^{t} y_i. \tag{3}$$

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 with forecasting purposes have been the most popular forecasting methods used in business and industry [13, 14]. Basically, the Single Exponential Smoothing (SES) consists of adjusting the previous forecast by weighting the forecast error, i.e:

$$F_{t+1} = F_t + \alpha(y_t - F_t), \tag{4}$$

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. However, if a reasonable number of observations are available, α can be estimated by minimizing the sum of the one-step-ahead squared forecast errors.

3.3. AR and ARIMA processes

[15] 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:

$$\phi(B)(1-B)^d y_t = \theta(B)a_t, \tag{5}$$

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, such as:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \tag{6}$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q. \tag{7}$$

The automatic identification procedure consists of selecting the best ARIMA model from a full range of possibilities according to the Akaike Information Criterion (AIC) normalized by sample size n that combines the fit of the model with the number of parameters used in order to avoid over parameterisation, i.e.

$$AIC_{p,q} \approx ln(\hat{\sigma}_a^2) + r\frac{2}{n}, \tag{8}$$

where $r = p + q$. The models estimated include orders: i) p=1,2,3; ii) q=1,2,3; and iii) d=1,2.

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:

$$y_t + a_1 y_{t-1} + \ldots + a_{n_a} y_{t-n_a} = b_1 u_{t-n_k} + \ldots + b_{n_b} u_{t-n_k-n_b+1} + a_t, \quad (9)$$

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 [16], and 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. The most widely employed architecture is the multilayer perceptron (MLP). 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 [17, 18].

Feed-forward architectures of MLPs are used to model nonlinear autoregressive NAR(p)-processes or NARX(p)-processes using exogeneous variables. 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

$$F_{t+1} = \beta_0 + \sum_{h=1}^{H} \beta_h g \left(\gamma_{0i} + \sum_{i=1}^{I} \gamma_{hi} p_i \right). \quad (10)$$

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 layer respectively. The β_0 and γ_{0i} are the biases of each neuron. I and H are the number of input and hidden nodes in the network and $g(\cdot)$ is a non-linear transfer function, which is usually either the sigmoid logistic or the hyperbolic tangent function ([19]). 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 [17].

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 2. 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:

$$F_{t+1} = \beta_0 + \sum_{h=1}^{H} \left(\beta_h g \left(\gamma_{0i} + \sum_{i=1}^{I} \gamma_{hi} p_i \right) \right) + \sum_{i=1}^{I} \delta_i p_i, \tag{11}$$

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.

All networks use for their training Bayesian Regularisation and no validation set is needed as in typical NN training [20]; 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:

$$f(p) = 1/(1 + exp^{-p}), \tag{12}$$

where p is an input. The networks are built using the Neural Network toolbox ($MATLAB^{TM}$) 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. All forecasts considered are one-step-ahead. 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. Obviously, the MdAPE is a more robust implementation of MAPE. 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 3 shows the Mean of the MAPE and MdAPE of the considered methods. In this table the minimum values 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 3, 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 6. We provide the mean error on the same figure as well. Again, the multivariate models show a better performance in comparison to the rest of the techniques.

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; therefore 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 from another method has no significant differences. The reader is pointed for more information to [21]. The numerical results of the non-parametric tests are provided in Table 4, whereas a visualisation of the results of the Nemenyi test is provided in Figure 7.

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 [22]. The test we use is based on:

$$y_t = \alpha_0 + \alpha_1 F_{1t} + \alpha_2 F_{2t} + e_t, \tag{13}$$

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 (13) can be examined either in an unconstrained or a constrained form, where in the latter $\alpha_1 + \alpha_2 = 1$. Here we use the later, since without the constraint the results show little more than the possible collinearity of the methods.

Table 5 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 5 we can conclude that all ARX, NARX and NARX-Lin contribute significantly to the univariate models. On the other hand, only NAR, MA and the Naïve methods contribute to ARX, but not to NARX or NARX-Lin. This can be explained by possible nonlinearities in the time series that are captured by all NAR, NARX and NARX-Lin methods, but not by ARX. Furthermore, we can observe that the inclusion of the direct linear modelling of information with the NARX-Lin method has a minor effect on the significant contribution of ARX on its nonlinear counterparts. From these results we conclude that the univariate methods have not captured additional significant information over the multivariate, whereas the opposite is true, due to access of the multivariate models to downstream information of the supply chain.

5. Discussion

We have provided evidence that information sharing can lead to significant improvements in forecasting accuracy. In this section we will investigate if there is any connection between the BWR and the accuracy gains of using downstream information of the supply chain. In other words we will try to establish the conditions under which information sharing can be beneficial in reducing forecasting error, considering the magnitude of bullwhip observed for each SKU.

The observed forecasting error for a particular SKU can be viewed as the sum of several factors that make the time series harder to forecast and more random. In general, the Bullwhip effect is thought to be one of these factors, making the forecast less accurate as one moves up-stream in the supply chain [2]. However, the nature and the magnitude of this connection is neither well established, nor well researched. Here we propose a simple framework to measure this.

Discounting the Bullwhip effect, the remaining error can be attributed to the forecastability of a product, i.e. the difficulty of forecasting a time series because of its properties and structure, i.e. presence of trend, seasonality, irregular components in the time series and so on. There are time series that are easy to forecast, while others can be more challenging, due to their statistical properties. Before we measure the impact of Bullwhip effect on an SKU, we need to determine its degree of forecastability first and quantify it. We propose to construct such a measure by calculating the relative forecasting error in different levels of the supply chain. More specifically in our case we are considering the levels of supplier and retailer, therefore we define as a degree of forecastability: $Relative\ Error = Supplier\ Error/Retailer\ Error$. Given the error measures we have already used this becomes:

$$\begin{aligned} Relative\ MAPE &= MAPE_{supplier}/MAPE_{retailer}, \\ Relative\ MdAPE &= MdAPE_{supplier}/MdAPE_{retailer}. \end{aligned} \tag{14}$$

Note that the proposed relative error includes potential forecasting errors due to the Bullwhip effect in the supplier time series, or more generally in the up-stream data. Therefore, although it does not allow us to separate the two sources of error, it allows us discount the internal forecastability of the time series. A ratio equal to 1 would imply that the difficulty of forecasting a particular SKU is equal, i.e. has the same forecasting errors, at supplier and retailer level. If the ratio is above 1 then the product is more difficult at supplier level, while the opposite would imply that forecasting that product becomes harder at retailer level. Therefore, once we calculate the relative errors at different levels of the supply chain we can investigate whether the change in forecastability is due to the Bullwhip effect or not. It is imperative that we consider the same forecasting method for producing forecasts at both levels of the supply chain, so that we have a fair comparison and not bias our errors in favour of a particular method. That implies that we have to use an adequate univariate method, as there is no downstream information sharing at the retailer level. Based on our forecasting accuracy results in table 3 we use SES to construct the relative accuracy ratio, given its good performance in producing forecasts at a supplier level. Table 6 provides the descriptive statistics of both relative errors defined in (14). On average there is a decrease in forecastability (increase of the forecasting errors) for the supplier of 360% and 298% according to relative MAPE and relative MdAPE respectively. It is noteworthy that while relative MAPE is always larger than 1, indicating a degredation of forecastability for the supplier time series, this is not always true for the relative MdAPE.

In section 2 we established a robust measurement for the BWR. We will use this as an estimate of the Bullwhip effect and investigate whether we can identify significant relationships between the degradation of forecastability and the BWR. Our hypothesis is that there should be a positive relationship between these two variables, which we test using regression modelling. We test both a linear and a quadratic specification of the model, to evaluate possible nonlinearities. These models are:

$$\begin{aligned} Linear: \ Relative\ Error &= \alpha_0 + \alpha_1 BWR + e, & (15) \\ Quadratic: \ Relative\ Error &= \alpha_0 + \alpha_1 BWR + \alpha_2 BWR^2 + e. & (16) \end{aligned}$$

A significant α_1 indicates that there is a linear connection between the degradation of accuracy and the BWR, with a positive coefficient implying that high BWR results in higher increase of the forecasting errors. Table 7 provides the estimated coefficients, their p-values for each case, the coefficient of determination R^2 and the adjusted R^2 for each model. We can see that there is a significant positive linear relationship in both MAPE and MdAPE, implying that indeed high BWR has a worse impact on forecasting accuracy. However, modelling BWR seems to account only for about 20% to 23% of the observed variability, depending on the error metric used. This becomes clearer in figure 8 where the relative errors are plotted against BWR. Furthermore, regarding relative MAPE there seems to be some evidence of nonlinearity that is not captured adequately by the quadratic model.

Instead of using the relative forecasting error we can model directly the supplier's error as a function of BWR and the retailer's error. Here we try to explicitly separate the forecasting error sources due to the time series properties and due to Bullwhip effect. In this case the models become:

$$\begin{aligned} Linear: \ error_{supplier} &= \alpha_0 + \alpha_1 error_{retailer} + \alpha_2 BWR + e, & (17) \\ Quadratic: \ error_{supplier} &= \alpha_0 + \alpha_1 error_{retailer} + \alpha_2 BWR + \\ & \quad \alpha_3 error_{retailer} \cdot BWR + \alpha_4 error_{retailer}^2 + \\ & \quad \alpha_5 BWR^2 + e, & (18) \end{aligned}$$

where $error_{supplier}$ and $error_{retailer}$ are the supplier and retailer forecast errors. The quadratic model includes the interaction term $error_{retailer} \cdot BWR$. The objective of these models is to evaluate whether with this formulation we can explain further $error_{supplier}$. Table 8 lists the coefficients, the p-values of the models, coefficient of determination R^2 and the adjusted R^2 for each model. In this case we can see that the quadratic model with the interaction term increases R^2 substantially, accounting for 47% or 40% of the variability of supplier's errors for MAPE and MdAPE respectively. Considering the model coefficients, the results for the linear models are similar to the previous models. There is a significant positive relationship of both $error_{retailer}$ and BWR with $error_{supplier}$, which means that high BWR reduces forecasting accuracy for the supplier. The quadratic models are more difficult to interpret, however we can see that there are significant nonlinearities, expressed either as quadratic or interaction terms, implying that there is potential for building nonlinear models to explain how the BWR affects forecasting accuracy.

Finally, we investigate whether the change in forecasting error due to the use of downstream information in the supplier forecasts is related to BWR. For this we calculate the error change for both MAPE and MdAPE as:

$$MAPE_{change} = (MAPE_{SES} - MAPE_{NARX-Lin})/MAPE_{SES}, \quad (19)$$
$$MdAPE_{change} = (MdAPE_{SES} - MdAPE_{NARX-Lin})/MdAPE_{SES} \quad (20)$$

where the subscripts SES and $NARX-Lin$ indicate the errors due to the respective models that demonstrated good performance in table 3. The correlation between $MAPE_{change}$ and BWR is -0.214, while for $MdAPE_{change}$ and BWR it is 0.101. Both are insignificant at 1%, 5% and 10% significance level. We provide scatterplots between the error change and BWR in figure 9. Fitting linear regression verifies that BWR is not a significant regressor and therefore we conclude that the forecasting error reduction due to information sharing is not explained by BWR. We found similar results for all multivariate models in this analysis.

6. Conclusions

The utility of information sharing with regards to forecasting performance is a controversial point. Theoretical analysis relying on restrictive assumptions claims that all 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% lower 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. In addition, no support was found for the assumption of some earlier work that the supplier can recover the statistical structure of downstream demand with no information sharing.

We also explored the connection between the extend of the Bullwhip effect, measured by the Bullwhip Ratio and forecasting accuracy. Initially we proposed a more robust measurement of BWR and showed that there is a significant positive relationship, i.e. higher BWR leads to higher forecasting errors for the supplier and we provided evidence of nonlinearities in this relationship. Once we consider the change in forecasting accuracy achieved by information sharing, we found no evidence that BWR can explain the observed error reductions. Therefore, while the BWR arises in part from downstream demand variability and supply chain processes, it does not determine supply forecasting accuracy once information sharing is considered.

Further research can be addressed on extending the case study to different industries and relaxing the mathematical restrictions to make wider the application of the results. Defining new measures for the Bullwhip Effect remains an open question and could depend on the company, industry or product category, etc. Is the same amplification observed for every category?

Different industries lead to different BE factors? More informative measures could also lead to better connection between upstream forecasting accuracy and the Bullwhip Effect. Then the gap between reality and theory may be reduced.

Acknowledgment

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

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] E. S. Gardner, Exponential smoothing: The state of the artpart ii, International Journal of Forecasting 22 (2006) 637–666.

[15] G. M. J. Box, G. C. Reinsel, Time Series Analysis Forecasting and Control, Prentice-Hall, Englewood Cliffs, 1994.

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

[17] 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–.

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

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

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

arXiv:http://www.mitpressjournals.org/doi/pdf/10.1162/neco.1992.4.3.415, doi:10.1162/neco.1992.4.3.415.
URL `http://www.mitpressjournals.org/doi/abs/10.1162/neco.1992.4.3.415`

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

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

FIGURES

List of Figures

1 Flow of material (-) and information (- -) for a two serially linked supply chain. 21
2 Example of a typical SKU. Retailer sales are in a solid line (-) and Volume received in a dashed line (- -) 22
3 Histogram of bullwhip ratio (BWR) calculated by the proposed robust and the conventional method. A bias towards smaller BWR is evident for the conventional BWR calculation. 23
4 Scatter plot between retailer and supplier mean sales. 24
5 Scatter plot between retailer and supplier sales standard deviation. 25
6 MAPE and MdAPE boxplots for forecasting methods. The thick blue line represents the mean of each error distribution. . 26
7 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. 27
8 Linear and quadratic fit of BWR on relative forecasting error. The dashed line is the linear fit, while the dotted line is the quadratic fit. Any point under the thick red dashed line implies that the forecasting accuracy upstream is higher than downstream, i.e. a relative error lower than 1. The equations for the linear fit are provided. 28
9 Linear fit of BWR on the percentage change of forecasting error caused by using downstream information. Any point over the thick red dashed line implies that the forecasting accuracy was improved. The equations for the linear fit are provided. . 29

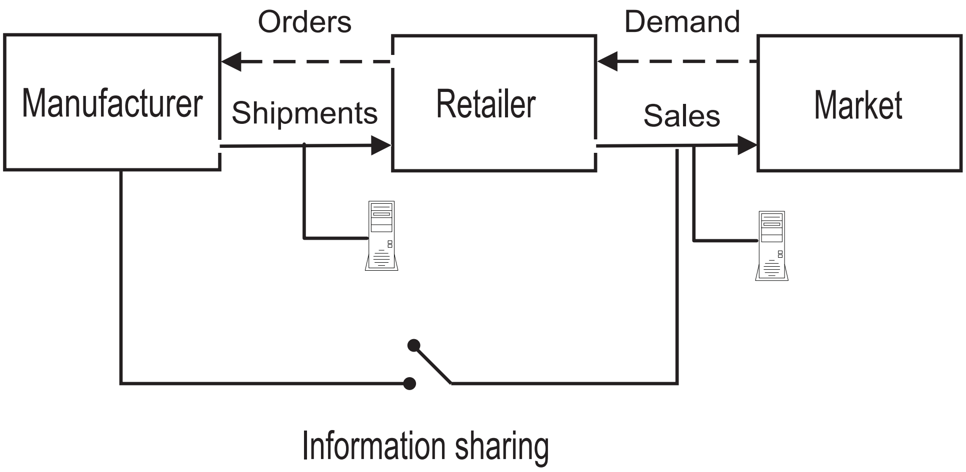

Figure 1: Flow of material (-) and information (- -) for a two serially linked supply chain.

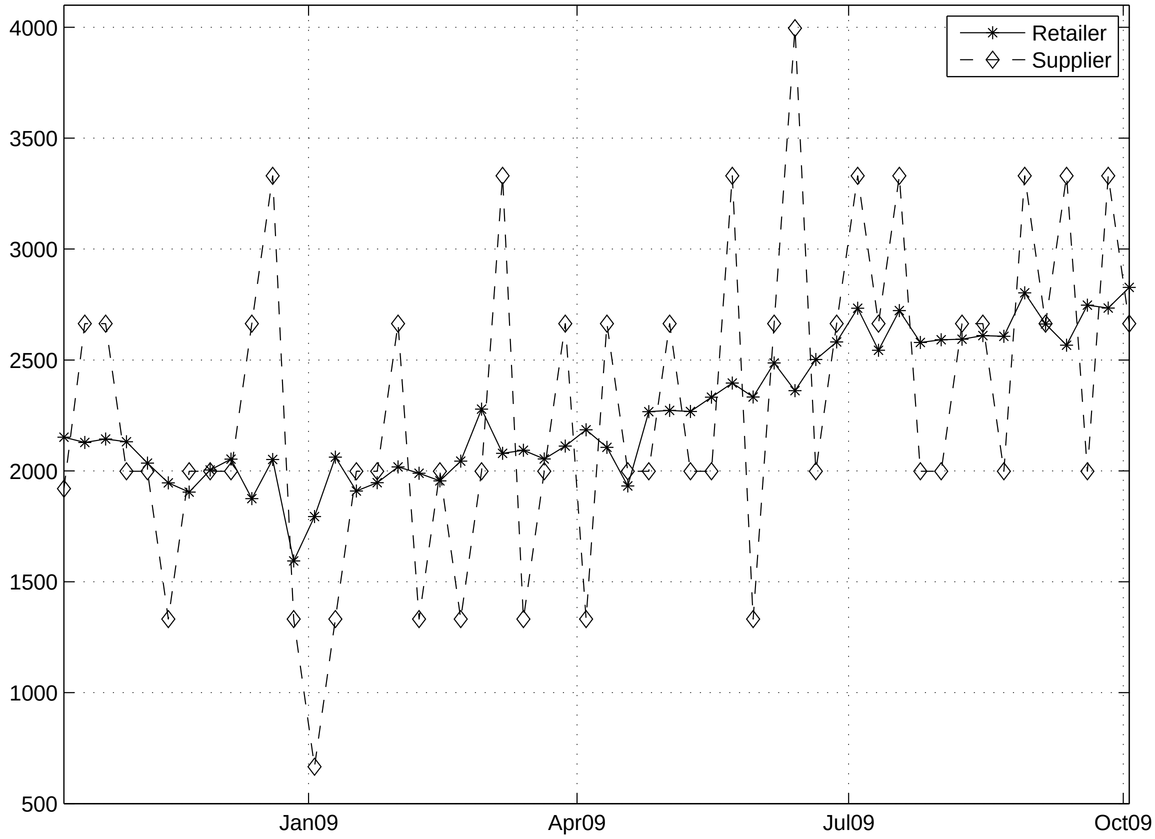

Figure 2: Example of a typical SKU. Retailer sales are in a solid line (-) and Volume received in a dashed line (- -)

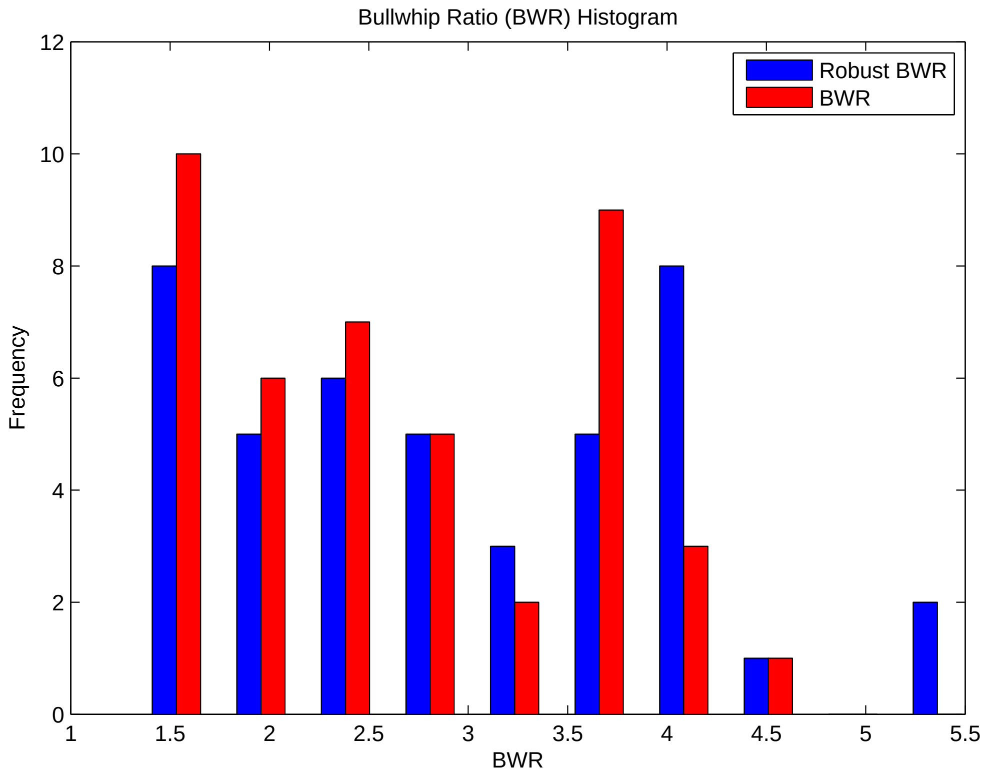

Figure 3: Histogram of bullwhip ratio (BWR) calculated by the proposed robust and the conventional method. A bias towards smaller BWR is evident for the conventional BWR calculation.

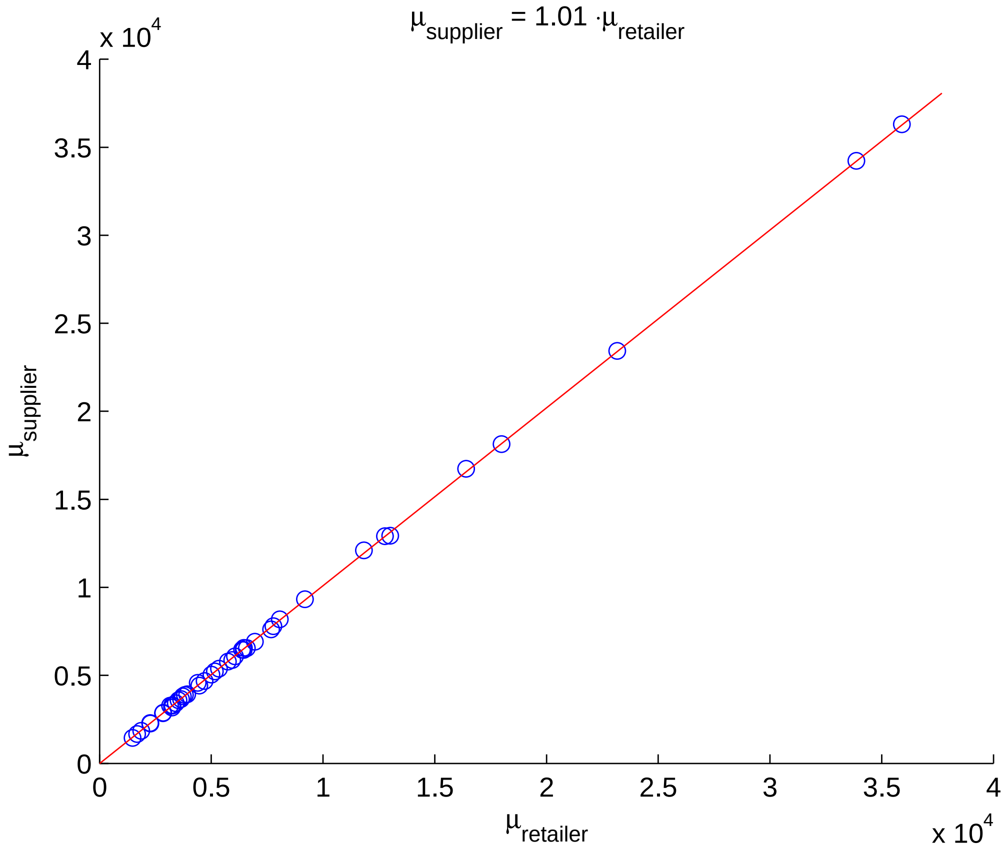

Figure 4: Scatter plot between retailer and supplier mean sales.

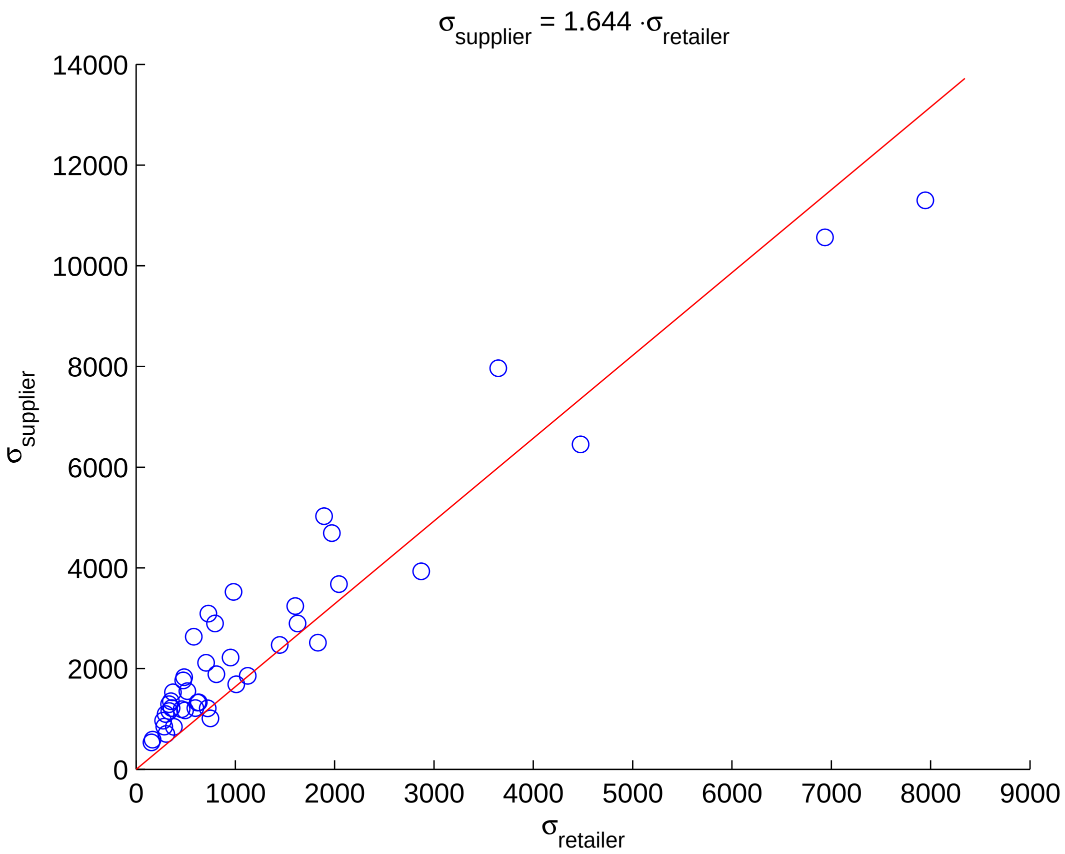

Figure 5: Scatter plot between retailer and supplier sales standard deviation.

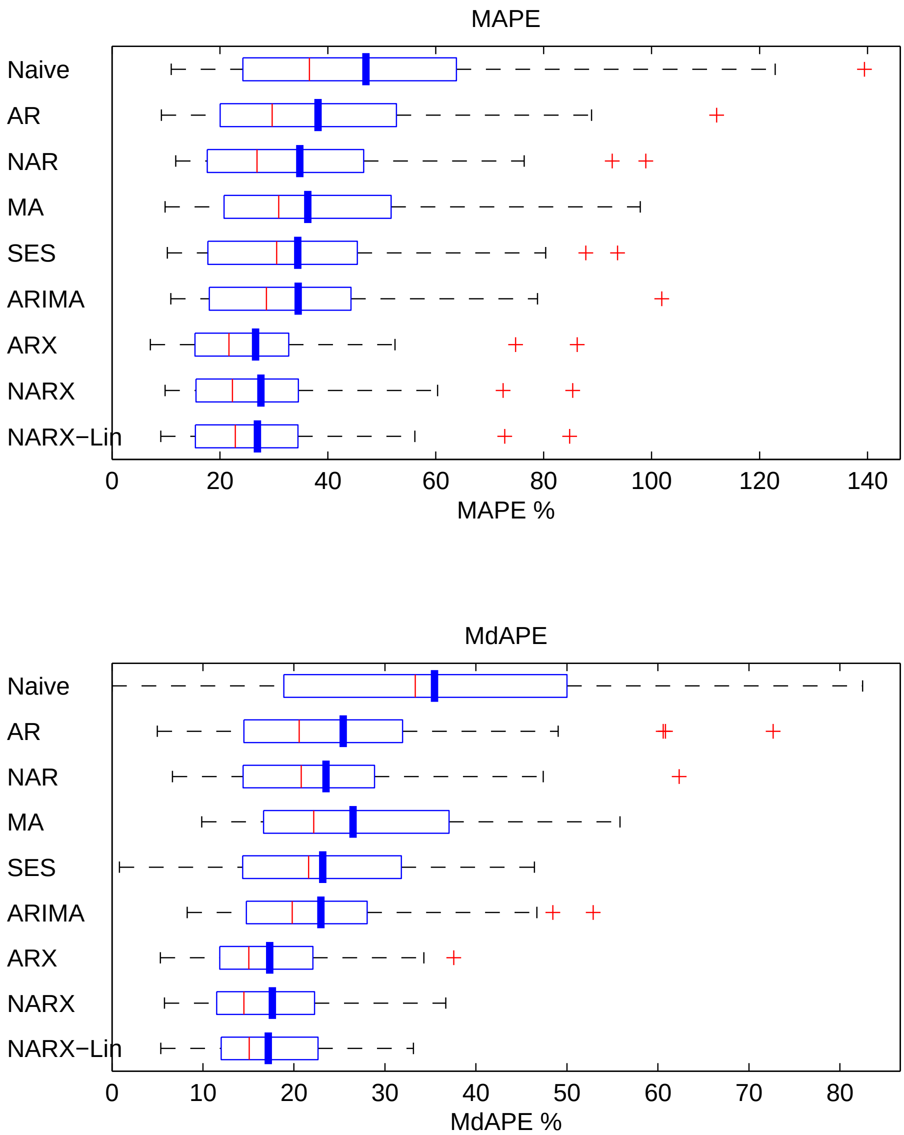

Figure 6: MAPE and MdAPE boxplots for forecasting methods. The thick blue line represents the mean of each error distribution.

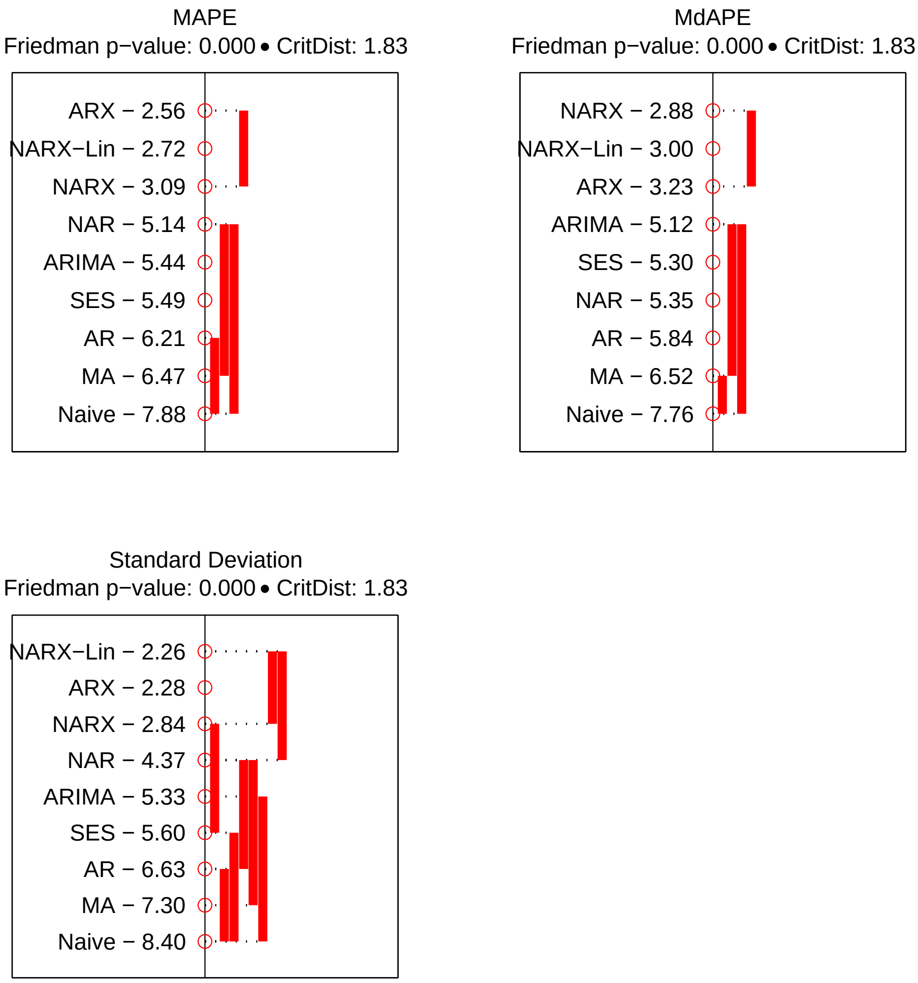

Figure 7: Methods are sorted from best (top) to worst (bottom). For methods joined by a red line there is no significant evidence of accuracy difference at 5% level. Multiple lines correspond to using different forecasting methods for evaluating the Nemenyi's test critical distance centre.

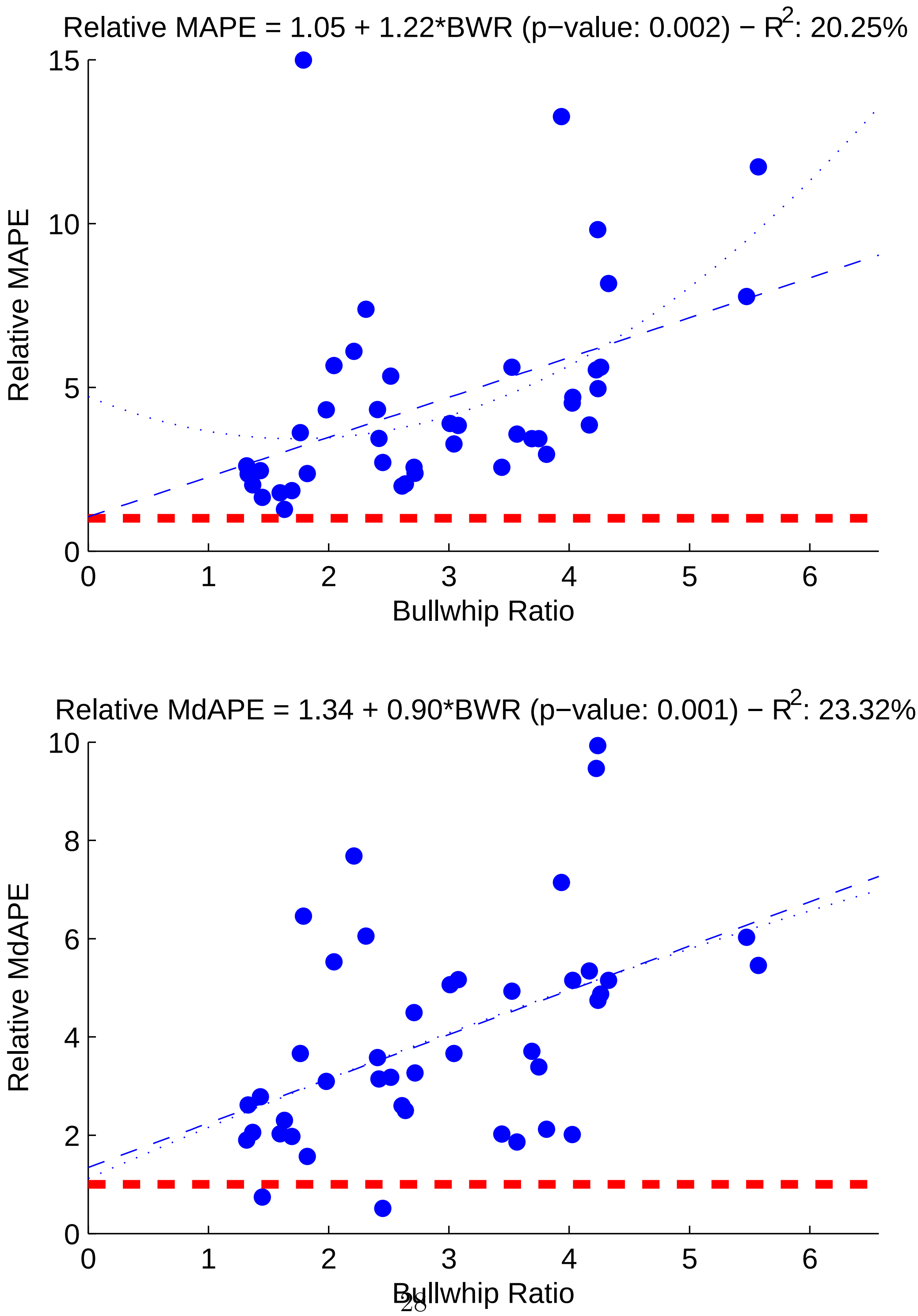

Figure 8: Linear and quadratic fit of BWR on relative forecasting error. The dashed line is the linear fit, while the dotted line is the quadratic fit. Any point under the thick red dashed line implies that the forecasting accuracy upstream is higher than downstream, i.e. a relative error lower than 1. The equations for the linear fit are provided.

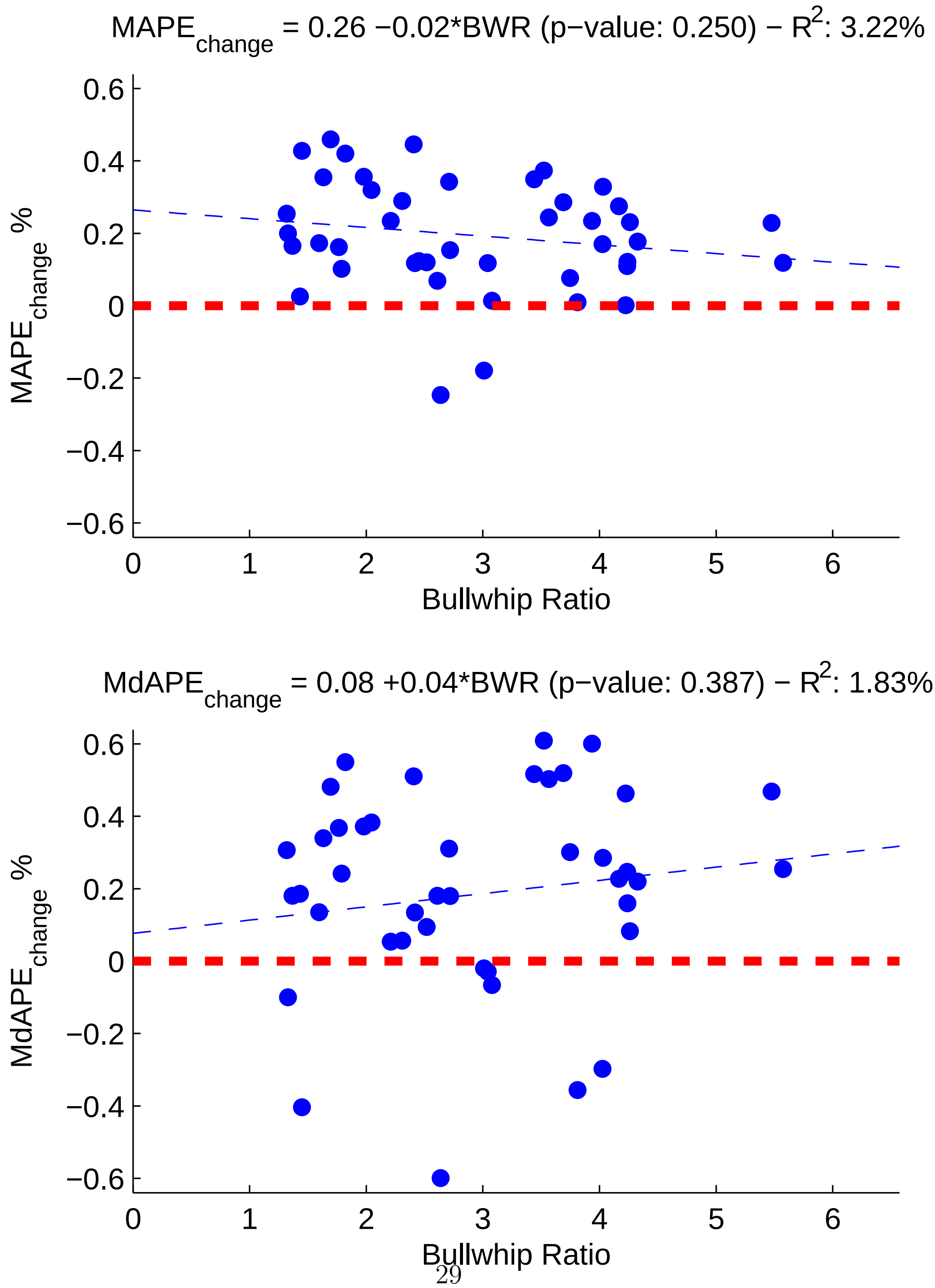

Figure 9: Linear fit of BWR on the percentage change of forecasting error caused by using downstream information. Any point over the thick red dashed line implies that the forecasting accuracy was improved. The equations for the linear fit are provided.

TABLES

List of Tables

1 Descriptive statistics for BWR. 31
2 Neural Network models design parameters. 32
3 Mean of the MAPE %, MdAPE % and standard deviation of the residuals for all forecasting methods. 33
4 Friedman and Nemenyi tests results. 34
5 Encompassing tests results. 35
6 Descriptive Statistics for Relative Errors. 36
7 Coefficients and R^2 of models using BWR to explain relative errors. 37
8 Coefficients and R^2 of models using downstream prediction error and BWR to explain upstream error. 38

Table 1: Descriptive statistics for BWR.

		Obs.	Min	Median	Mean	Max	Std
	Low	25	1.34	2.02	1.94	2.65	0.41
BWR	High	18	2.97	3.61	3.63	4.54	0.42
	All	43	1.34	2.37	2.65	4.54	0.94
	Low	26	1.32	2.13	2.13	3.08	0.57
Robust BWR	High	17	3.44	4.03	4.13	5.57	0.59
	All	43	1.32	2.71	2.92	5.57	1.14

Table 2: 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 3: 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 4: 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 5: 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.

Table 6: Descriptive Statistics for Relative Errors.

	Obs.	Min	Median	Mean	Max	Std
Rel. MAPE	43	1.28	3.62	4.60	14.99	3.09
Rel. MdAPE	43	0.51	3.58	3.98	9.93	2.14

Table 7: Coefficients and R^2 of models using BWR to explain relative errors.

		Linear	Quadratic
MAPE	α_0	1.05 (0.381)	4.72 (0.111)
	α_1	**1.22 (0.003)**	-1.49 (0.457)
	α_2	-	0.43 (0.173)
	R^2	0.203	0.239
	adj. R^2	0.183	0.201
MdAPE	α_0	1.34 (0.100)	1.12 (0.580)
	α_1	**0.90 (0.001)**	1.07 (0.442)
	α_2	-	-0.03 (0.902)
	R^2	0.233	0.234
	adj. R^2	0.215	0.195

Number in brackets are p-values. Values in boldface are significant at 5% level.

Table 8: Coefficients and R^2 of models using downstream prediction error and BWR to explain upstream error.

		MAPE	MdAPE
Linear	α_0	0.01 (0.927)	0.07 (0.237)
	α_1	**1.64 (0.001)**	**1.05 (0.002)**
	α_2	**0.06 (0.028)**	**0.03 (0.060)**
	R^2	0.256	0.213
	adj. R^2	0.219	0.174
Quadratic	α_0	**0.67 (0.037)**	0.12 (0.434)
	α_1	-1.47 (0.630)	3.22 (0.096)
	α_2	**-0.35 (0.020)**	-0.07 (0.344)
	α_3	**1.63 (0.048)**	0.02 (0.960)
	α_4	-0.75 (0.882)	**-9.83 (0.036)**
	α_5	**0.05 (0.016)**	0.02 (0.145)
	R^2	0.470	0.404
	adj. R^2	0.398	0.323

Number in brackets are p-values. Values in boldface are significant at 5% level.