The value of point of sales information in upstream supply chain forecasting: an empirical investigation

Mahdi Abolghasemi[a], Bahman Rostami Tabar[b], Aris Syntetos[c]

a: School of Mathematics and Physics, The University of Queensland, Brisbane, QLD, Australia,

email: m.abolghasemi@uq.edu.au

b: Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, United Kingdom,

email: Rostami-TabarB@cardiff.ac.uk

c: Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, United Kingdom,

email: SyntetosA@cardiff.ac.uk

Abstract

Traditionally, manufacturers use past orders (received from some downstream supply chain level) to forecast future ones, before turning such forecasts into appropriate inventory and production optimization decisions. With recent advances in information sharing technologies, upstream supply chain (SC) companies may have access to downstream point of sales (POS) data. Such data can be used as an alternative source of information for forecasting. There are a few studies that investigate the benefits of using orders versus POS data in upstream SC forecasting; the results are mixed and empirical evidence is lacking, particularly in the context of multi-echelon SCs and in the presence of promotions. We investigate an actual three-echelon SC with 684 series where the manufacturer aims to forecast orders received from distribution centers (DCs) using either aggregated POS data at DC level or historical orders received from the DCs. Our results show that the order-based methods outperform the POS-based ones by 6%-15%. We find that low values of mean, variance, non-linearity and entropy of POS data, and promotion presence negatively impact the performance of the POS-based forecasts. Such findings are useful for determining the appropriate source of data and the impact of series characteristics for order forecasting in SCs.

Information sharing, Supply chain forecasting, POS data, Promotions, Time series characteristics.

1

1 Introduction

In recent decades, different collaborative strategies have been developed to reach potential synergies across the supply chain (SC). There are various studies that discuss different types of SCs, from those exhibiting no collaboration at all, to those associated with collaborative arrangements to enhance forecasting and inventory performance. (Singh et al. 2018, Chan & Chan 2009, Lau et al. 2004). Such collaborative strategies include, but are not limited to, quick response (QR) (Perry & Sohal 2001), vendor managed inventory (VMI) (Choudhary & Shankar 2015), collaborative planning, forecasting, and replenishment (CPFR) (Aviv 2001), and point of sales (POS) information sharing (Williams & Waller 2011, Narayanan et al. 2019).

In this paper, we focus on the value of POS information sharing to improve forecasting performance in SCs. Traditionally, historical orders have been the main source of information for upstream forecasting. More recently, technological advancements have enabled companies to record and share POS data as an expedient source of additional information in SCs. Sharing POS data, in particular, has been put forward as an important mechanism to increase downstream visibility and may be utilized for various purposes such as demand forecasting. Downstream information should be shared with upstream members as much as possible. This can increase visibility across a SC and reduce the associated uncertainty which ultimately can result in more accurate forecasts (Christopher & Holweg 2011).

POS information can be used for forecasting across different levels of the SCs. Retailers can use them directly to forecast customers' demand. Sharing POS information might be valuable to forecast retailers' demand in downstream SCs (Narayanan et al. 2019, Hartzel & Wood 2017). POS data, if shared with the upstream SCs, can be used by manufacturers to forecast DCs' orders. There are several studies that compare the forecasting accuracy of the POS- and order-based methods in forecasting retailers' and DCs' orders. While some studies advocate the use of order-based forecasts for DCs' orders (Narayanan et al. 2019, Williams & Waller 2010), others have found the POS-based forecasts to be more insightful for decision making (Croson & Donohue 2003). The literature suggests that various factors such as demand pattern, product type, and inventory

policy, may impact the value of POS and order data for forecasting. These factors can be classified into two main categories: i) operational factors, and ii) time series factors.

The first category, operational factors, refers to the products and operational activities of the SC members. This includes product characteristics such as product types and product family, and operational activities and constraints of SC stakeholders such as inventory policy and ordering system (Williams & Waller 2010). The second category refers to the characteristics of the POS and order time series, e.g., demands and orders distribution, trend, seasonality, volumes, promotion status, presence or absence of bullwhip effect, and volatility (Croson & Donohue 2003, Williams et al. 2014, Hartzel & Wood 2017, Narayanan et al. 2019).

These studies have some limitations which are summarised as follows: i) the impact of time series characteristics on the relative value of POS and order series have not been investigated thoroughly in the presence of promotions ii) the canonical features of POS data and orders including trend, seasonality strength, non-linearity, entropy, hurst (self-similarity or long dependency), stability, skewness, and kurtosis and their impact on forecasting accuracy have been largely ignored iii) the focus of research studies has been on investigating the value of POS data at downstream levels of SCs, i.e., looking at whether using POS data can help retailers to forecast their demand more accurately and decrease their inventory level. However, POS data can be aggregated and shared with the upstream SC members for various purposes such as operations management and strategic planning. The value of POS data in upstream SC has not been investigated enough. iv) there is no rigorous empirical study with a sufficiently large dataset to determine the impact of POS and order series characteristics on forecasting performance. Therefore, there is a need for further empirical studies with large datasets to better understand the value of POS data for upstream forecasting to allow managers to determine whether such information should be used and if so under what circumstances. Our study is motivated by a real-world problem and real data collected from 684 stock keeping unit locations (SKULs), making it the largest case study among others.

We address the above-discussed gaps by considering a major food company producing fast-moving consumer goods (FMCG) in Australia. Generating accurate demand and order forecasts in the food

FMCG helps to minimise stockouts, reduce excess inventory, and improve availability to customers which all positively impact competitiveness and profitability in organisations (Lau et al. 2004). In order to forecast the manufacturer's demand, we use four different forecasting methods -Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing in state-space form (ETS), Regression with ARIMA errors (R-ARIMA), and ARIMA with explanatory variable (ARIMAX)- and feed them with i) the shared POS data that is aggregated at DCs level, and ii) historical orders of DCs, individually. We use ARIMAX and R-ARIMA methods as our main forecasting models as they take into account external information to generate forecasts and have shown promising results in different forecasting problems (Abolghasemi, Beh, Tarr & Gerlach 2020). The ARIMA and ETS forecasting methods are popular in academia and practice and serve as basic benchmark methods. We also benchmark the predicted demand against the company's forecasts. This is the natural benchmark recognizing though that the main intention here is to derive insights into the comparative performance between POS- and order-based forecasts, rather than making sure that we do better than the benchmarks. We then analyze ten canonical characteristics of the orders and POS series that can describe the series behavior including mean, variance, trend, seasonality strength, non-linearity, hurst, stability, entropy, skewness, and kurtosis to determine their impact on the comparative performance of POS- and order-based series and specify the conditions under which the POS or order series may be more suitable for forecasting purposes. We should note that while it is possible to combine several sources of information such as POS and orders to generate the forecasts, in this study, we investigate which source of information can improve the forecast accuracy rather than finding a possible method to combine POS and order information. Combining POS and order information has been studied for the pharmaceutical industry (Van Belle et al. 2020) but it requires a different study in the retailing context as combining POS and order information in the presence of promotions is a challenging task.

We summarize our contributions to the literature as follows:

1. We investigate the value of POS information to forecast DCs' orders in a SC in the presence of promotions by means of an empirical study in the FMCG industry.

4

2. We use aggregated POS data in upstream SC as opposed to downstream SC where POS data is directly accessible.

3. We examine features of POS and order time series including mean, variance, trend, seasonality strength, non-linearity, hurst, stability, entropy, skewness, and kurtosis and develop a hierarchical linear model (HLM) to determine which characteristics have a statistically significant effect on the relative performance of the POS-based to order-based forecasts.

The remainder of the paper is organized as follows. Section 2 reviews the literature. Section 3 introduces the case study and describes the dataset of our empirical analysis. Section 4 explains the experiment setup and Section 5 provides a thorough discussion of the obtained results. Finally, conclusions are drawn in Section 6.

2   Information sharing in supply chain

There are studies that show information sharing, in general, can benefit the SCs by reducing the bullwhip effect (Disney & Towill 2003, Wang et al. 2016), increasing service levels (Claassen et al. 2008), saving inventory (Lau et al. 2004), and reducing cost (Lau et al. 2004, Claassen et al. 2008). Despite the benefits of information sharing in the SCs, according to two surveys conducted by Capgemini and Forrester Research, only 40% and 27% of retailers shared data with their suppliers, respectively (Seifert 2003). Various reasons are reported as obstacles for information sharing such as organizational trust, benefits and cost of information sharing, commitment, and cultural issues (Mishra et al. 2007). In the particular case of sharing POS information, the inability to effectively use the POS data has been a major obstacle for sharing POS information (Barratt 2003, Alftan et al. 2015, Bassamboo et al. 2017).

Studies suggest that both operational factors and demand characteristics, e.g., product type, demands and order size, frequency of orders, the variance of orders, and the level of required forecasts in SCs (downstream or upstream), impact the value of POS information for forecasting in SCs (Hartzel & Wood 2017, Narayanan et al. 2019). The impact of operational factors on the forecasting

accuracy of POS- and order-based approaches has been investigated extensively. For example, Hosoda et al. (2008) found that using POS data in a two-echelon SCs, which has order-up-to ordering policy and ARIMA demand, decreases the standard deviation of manufacturer's demand forecasting error by between 8% and 19%, and, consequently, reduces inventory-related costs. Product types and their demand are commonly investigated as a factor that can impact the value of POS data in forecasting. Williams & Waller (2010) investigated 24 products from three different categories (cereal, yogurt, canned soup) in 18 regional DCs in the USA. They forecasted DCs' orders with naive, moving average, exponential smoothing, and Holt-Winters methods and found that the POS-based forecasts significantly outperformed the order-based ones for cereal products. They also found that the POS-based forecasts is more accurate than the order-based forecasts in 65% of the observations. Their performance varied across different products with different characteristics. The POS-based forecasts tend to be less accurate for products with large values of bullwhip effect (large ratio of orders variance to POS variance) and high non-turn volume (large promotional sales).

In one of the few studies that examine the value of POS data for forecasting in upstream SCs, Narayanan et al. (2019) evaluated the performance of POS and order data to forecast orders at suppliers and DCs levels, and demand at all levels. They analyzed 21 products (from cereals, detergent, frozen pizza, and cosmetics) and found that the POS-based forecasts do not improve the accuracy of the order forecasts at DCs' level.

Time series characteristics is another family of factors that can play an important role in determining the forecasting accuracy of methods. Several studies have shown that time series with different features impact the performance of forecasting methods (Kang et al. 2017, Abolghasemi, Beh, Tarr & Gerlach 2020, Spiliotis et al. 2020, Abolghasemi et al. 2022). While there are many characteristics that can be extracted from time series (Fulcher & Jones 2017), a smaller number of important characteristics can be used to adequately represent the time series behavior (Kang et al. 2017). In a comprehensive study on time series characteristics, Wang et al. (2009) analyzed more than 300 datasets and extracted nine important characteristics of time series including trend, seasonality, periodicity (to determine the seasonality and cyclic patterns), serial correlation,

skewness, kurtosis, nonlinearity, hurst (hurst or self-similarity is an index that indicates the long-range dependency in time series), and chaos (chaos or entropy is indicative of time series forecastability which is defined by the level of signals in time series). They applied ARIMA, exponential smoothing, random walk, and neural network methods on their data and exploited the results in the form of six judgmental rules by which the most accurate forecasting method according to the time series characteristics were identified. For example, when time series have a strong trend, high non-linearity, hurst, low seasonality, low kurtosis and skewness, low chaos, and no noise, then the neural network is the most accurate method, followed by ARIMA, exponential smoothing, and random walk. Spiliotis et al. (2020) also used a small set of canonical characteristics to compare the M3 and M4 forecasting competitions datasets. They extracted the entropy, trend, seasonality, linearity, and stability as the primary characteristics of the M3 and M4 series and further exploited skewness, kurtosis, non-linearity, and hurst to describe and compare the behavior of their time series. Hartzel & Wood (2017) examined over 60,000 orders for more than one hundred different items across 25 DCs, and found that using the real-time POS data increases the forecasting accuracy of orders by 11.2%. They showed that using the POS data improves the accuracy of order forecasts under certain conditions: i) When orders have a low level of frequency, ii) When orders are relatively small or large, iii) When orders have moderate variability. They measured variability by the log of variability of item orders and ranged it on the scale of 0 to 100 percentiles to represent the strength of variability. They indicated the items between 40 to 60 percentiles with moderate variability. They did not provide any specific information about demand characteristics such as trend, seasonality, and promotion status.

We posit that the performance of POS- and order-based forecasts shall be analyzed according to the characteristics of the input series rather than merely their nature, i.e., being POS or order series. We argue that the fundamental reason that makes POS or order series more suitable for forecasting in SCs is not their nature (being POS or order), rather it is their characteristics that render them more suitable for a specific forecasting purpose.

When it comes to the value of sharing POS data in SCs, previous studies reveal some limitations

that are summarised as follows:

- The focus of the studies has been on the downstream SCs, mainly discussing the value of sharing POS data in forecasting retailers' orders. Few empirical studies have been conducted to investigate the value of shared POS data in upstream SCs (Narayanan et al. 2019, Williams & Waller 2010, 2011).

- The information-sharing literature has minimally considered the impact of POS and order series characteristics. While the impact of factors such as demand distribution, demand pattern, and products features have been studied, there is no rigorous empirical analysis to investigate the canonical characteristics of orders and POS time series that carry useful information about their behavior. This includes mean, variance, trend, seasonality strength, non-linearity, hurst, stability, entropy, kurtosis, and skewness. We also consider promotion as another important factor. We then evaluate the association between these characteristics and their impact on the accuracy of forecasts.

- There are mixed results about the benefits of using POS information for forecasting in SCs. While some studies found using POS information can increase the forecasting accuracy of the SCs orders at upstream levels (Williams & Waller 2010), other studies found that using POS is useful for demand planning at retailers' and suppliers' levels (Narayanan et al. 2019). More studies are required to investigate the impact of various factors on the forecasting accuracy of the POS-based methods and determine the conditions under which sharing POS data is beneficial for SCs forecasting (Williams et al. 2014).

3   Case and empirical data

We consider a three-echelon SC of FMCG in Australia with one manufacturer, twelve DCs, and more than 2000 retailer shops to investigate the value of POS information in upstream SC forecasting. While we gather POS information from retailers, our focus is on the upstream SC where

we aim to forecast the DCs' orders to the manufacturer. Figure 1 symbolically illustrates the information flow for our case study SC with one manufacturer, two DCs, and six retailers.

In our case study SC the retailers place orders to their corresponding DC and DCs place orders to the manufacturer. We look at this problem from the manufacturer's point of view. The manufacturer aims to forecast the DCs' orders (manufacturer's demand) for production planning, logistics, and inventory management purposes. The manufacturer does not have access to either DCs' operational information such as inventory levels and ordering policy or retailers' orders. Instead, the manufacturer has access to the aggregated-POS data at DCs level that is collected by a third party organization. The aggregated POS data corresponds to 12 DCs each of which includes data for several hundred retailer shops. There are in total 2000 retailers but we do not have the exact number of retailers that correspond to each DC. However, the scale of sales in each DC indicates that some of them have a larger market share than others. This data becomes available within a week. Therefore, the manufacturer has access to two sources of information to generate forecasts: i) historical orders of the DCs, and ii) the aggregated POS data at the DCs level. The manufacturer uses the former in a statistical model (Holt-Winters exponential smoothing) and utilizes the latter source of information to judgmentally adjust the output of the statistical model. Holt-Winters model is a relatively easy model to use and highly interpretable, making it a popular model for industries and our case study to use in practice Syntetos et al. (2016).

The forecast adjustment is performed by a team of experts who gather POS and market information from the retailers and use their experience to judgmentally adjust the output of the statistical forecast. The size of the adjustment is larger for promotional periods as the statistical model is not able to effectively incorporate promotional information. This is a cumbersome and cognitively demanding task as experts need to reconcile information and manually produce a large number of forecasts. This is particularly problematic when a promotion is offered as it requires experts to process a large amount of information which may introduce some bias to the forecasts Abolghasemi, Hurley, Eshragh & Fahimnia (2020). In addition, it is not evident which source of information is more suitable for forecasting. In order to automate the forecasting procedure, we use the POS
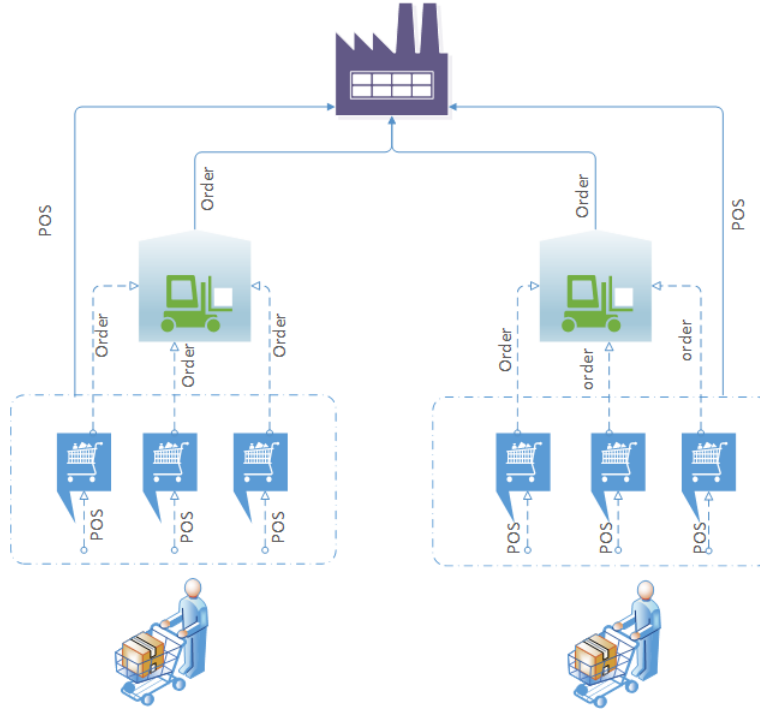
- Figure 1 Caption: Supply chain and the flow of information.

- Figure 1 Alt Text: The flow of information in a supply chain with one manufacturer, two main retailers and six DCs.

information in a quantitative model that is inspired by the case study problem and tailored to the current practice. We empirically investigate the comparative benefits of using POS and order series for forecasting purposes to help forecasters choose the right source of information and model for forecasting.

We conduct our analysis using data collected from the case study. Our data spans over 112 weeks between September 2016 to October 2018 and includes the weekly aggregated POS series, historical orders, promotion time, promotion type, and price for 684 SKULs. The SKUL refers to a stock keeping unit at different DCs. The SKULs are from different categories and families including breakfast cereals, spreads, and long-life dairy products but their label is unknown for us. These SKULs depict different levels of sales, orders, trends, seasonality strength, promotion frequency, and promotion type representing a diverse set of POS and order for analysis.

3.1 The characteristics of POS and order series

In order to understand how POS series are different from orders, we investigate their characteristics. While there are many characteristics that can be analyzed for each time series (Fulcher & Jones 2017), we exploit ten important characteristics of time series that can explain the behavior of the order and POS series. Identifying and analyzing these characteristics help managers to appreciate how consumers' demands evolve across SCs. These characteristics are defined in Appendix A. We use the '*tsfeatures*' package in R (Hyndman, Kang, Montero-Manso, Talagala, Wang & Yang 2018) to compute these characterises for POS and order series.

In order to gain insights on the comparative behavior of order and POS series, we measured the Kullback-Leibler divergence (KL) between order (target probability distribution) and POS series (Cover 1999). Although KL distance is asymmetric, it can measure how much information is lost when we approximate one distribution with another. Suppose $H(p)$ is the entropy of the target distribution, then we can measure the difference between the estimated and target distribution by $\frac{KL}{H(p)} * 100\%$. Table 1 provides the descriptive statistics of the order and POS series characteristics for first (Q1), second (Q2), and third (Q3) quartiles and the KL divergence (percentage differences in parenthesis), on overage. Figure 2 shows the histogram of the order and POS series for each characteristic and provides information about the distribution of these characteristics across all series.

In general, both POS and order series depict similar distribution (less than 5% difference) but different values in Q1, Q2, and Q3. Both POS and order series depict strong entropy. The series do not show strong trend, seasonality strength, non-linearity, hurst, and stability. POS and order series depict relatively small degrees of skewness and kurtosis which are largely caused by promotion. We investigated the pair of order and POS data for each product in our case study. Figure 3 shows the behavior of POS and order data for a typical product. We observe that the average quantity of order and POS series often vary around each other but they are significantly different during, before, and after promotional periods. Before promotional periods, orders tend to be larger than the POS series for that period, during promotional periods POS is extremely larger than order series, and
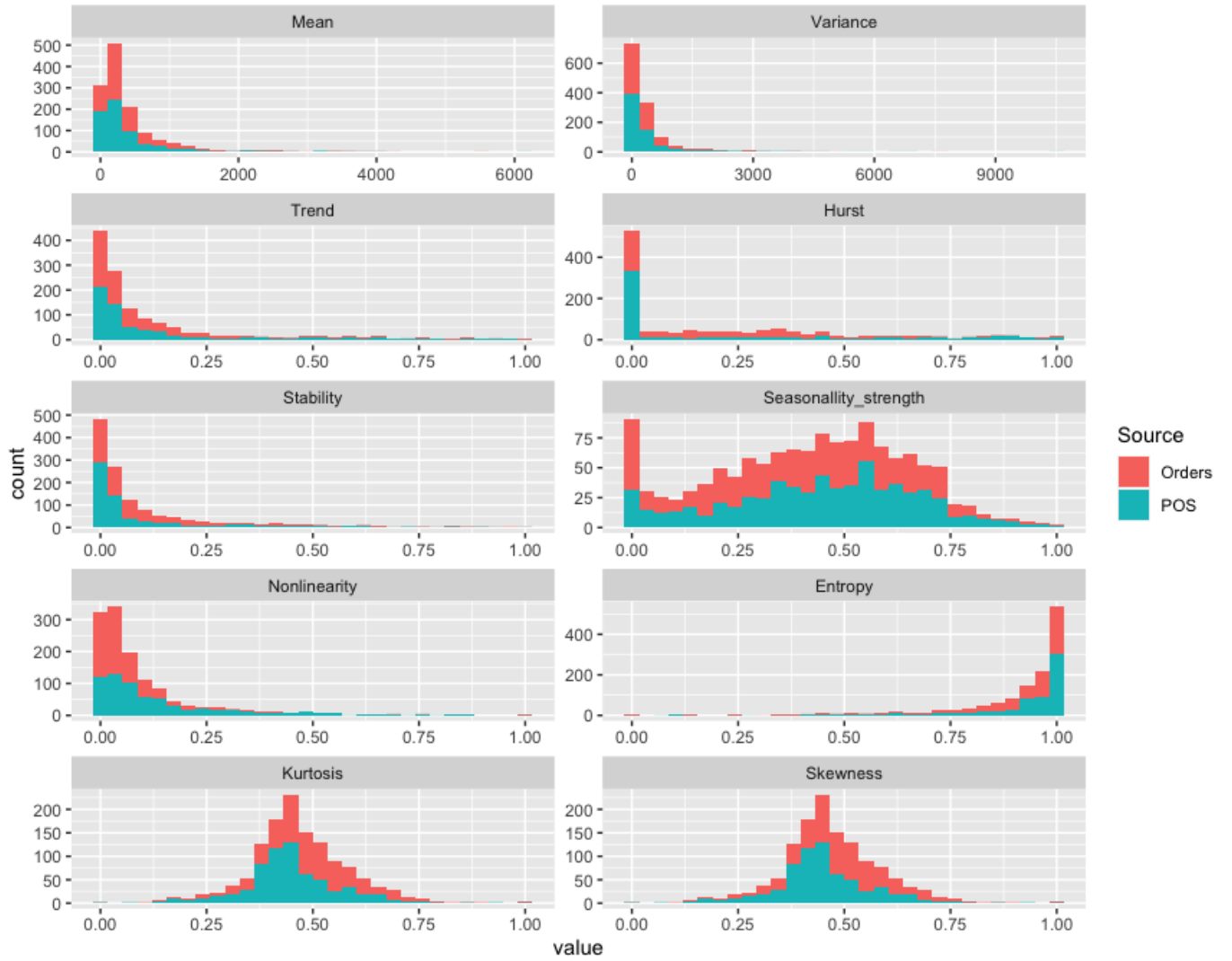
- Figure 2 Caption: Histogram of order and POS series characteristics.

- Figure 2 Alt Text: The distribution of main characteristics of POS and orders series across 684 SKULs.

Table 1: Descriptive statistics of order and POS characteristics

Characteristics	Orders			POS			KL-divergence
	Q1	Q2	Q3	Q1	Q2	Q3	Average
Mean	140.45	271.03	511.40	92.70	198.37	408.56	0.44 (4.23%)
Variance	83.55	174.91	425.77	48.86	120.69	330.55	0.42 (4.16%)
Trend	0.00	0.03	0.08	0.01	0.04	0.15	0.15 (0.90%)
Hurst	0.15	0.19	0.28	0.15	0.15	0.25	0.17 (2.69%)
Stability	0.01	0.05	0.15	0.01	0.03	0.17	0.06 (0.99%)
Seasonality strength	0.14	0.26	0.36	0.21	0.32	0.42	0.36 (3.62%)
Non-linearity	0.09	0.25	0.53	0.06	0.17	0.42	0.87 (3.98%)
Entropy	0.95	0.98	1.00	0.93	0.98	1.00	0.22 (3.19%)
Kurtosis	0.14	0.18	0.46	0.22	0.34	0.53	0.28 (2.76%)
Skewness	0.16	0.23	0.53	0.26	0.43	0.51	0.33 (2.99%)

after promotional periods often POS is larger than orders but this may vary from one product to another. The larger amount of orders before promotional periods is due to smoothing orders by DCs into a few periods before promotions, and the smaller size of orders after promotional periods indicate that often DCs have enough inventory after promotions and do not place new orders. DCs on average place larger orders to avoid the possibility of stock-out during promotional periods. Note that the smaller values of POS than orders could also be due to the possible stock-out in retailers' inventory but we do not have access to this information for verification.

Orders, on average, depict smaller values of the variance than the POS series as it is evident from Q1, Q2, and Q3 of variance for POS and order series. Figure 4 shows the histogram of the bullwhip effect for promoted and non-promoted products. We found that 489 SKULs out of 684 SKULs (71%) have higher variance in the orders than POS, i.e, they exhibit bullwhip effect phenomenon. More precisely, 196 out of 200 non-promoted series (98% of the non-promoted SKULs) and 293 out of 484 promoted series (60% of the promoted SKLUS) exhibit bullwhip effect phenomenon. There are 195
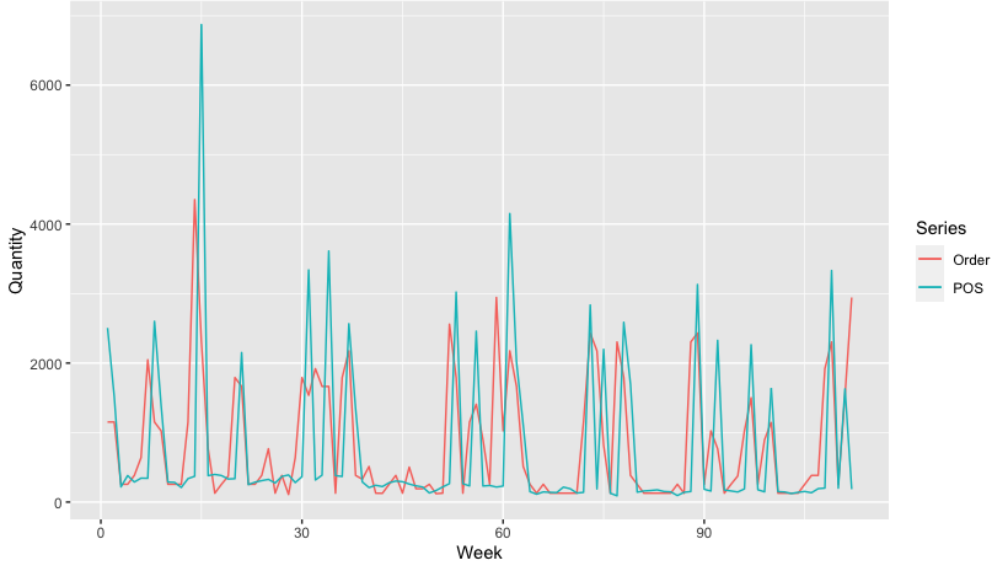
- Figure 3 Caption: Behaviour of orders and POS series for a typical product.

- Figure 3 Alt Text: The behaviour or orders and POS time series for one product is shown.

SKULs (29% of all SKULs) that showed anti-bullwhip effect behavior, i.e, the variability of orders series was smaller than POS series. The anti-bullwhip effect phenomenon often occurs during promotions when DCs deliberately smooth their orders during promotions. That is, DCs may place their promotional orders in two periods rather than at once. Therefore, orders show lower levels of variability than the POS series, and consequently the indicative fraction for bullwhip, i.e., $\frac{\sigma^2_{order}}{\sigma^2_{pos}}$, during promotional periods will decrease and contribute to anti-bullwhip behavior at some scenarios. This may have impacted their non-linearity and entropy as POS series depict a stronger non-linearity and entropy than the order series. Both order and POS series have shown similar values of hurst. In terms of seasonality, both POS and orders do not depict strong seasonality and POS series have slightly stronger seasonality. POS and order series have shown similar values of skewness and kurtosis with POS slightly having larger values. The series with large values of kurtosis and skewness are the ones that are heavily impacted by promotions.

In summary, we can see that POS and order series depict different levels of strength in each characteristic. However, our analysis shows that the distribution of the POS and order series are similar at 5% of significance level in all investigated characteristics. In Section 4, we present a

14
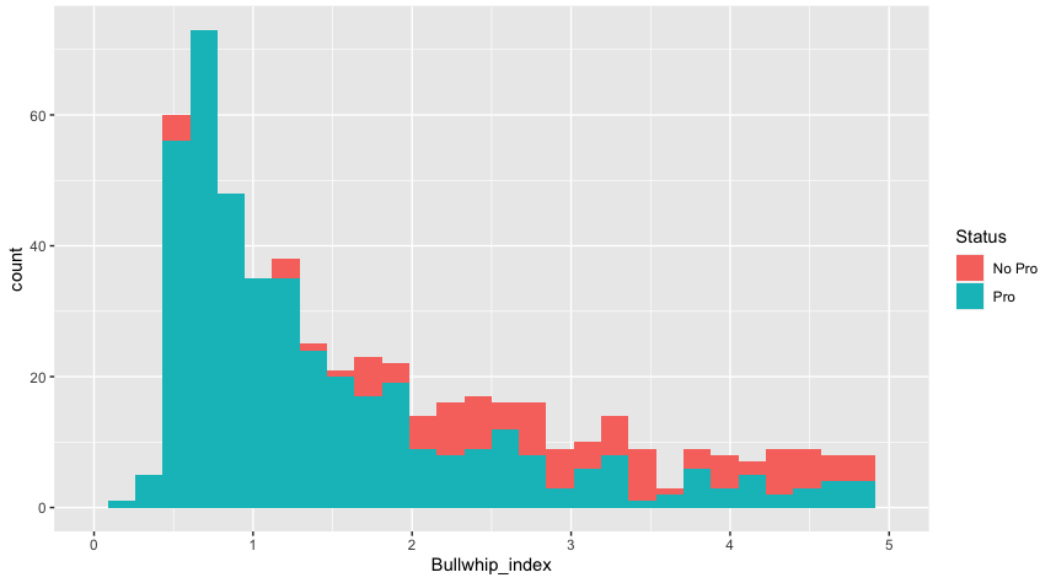
- Figure 4 Caption: The histogram of Bullwhip effect across all products.

- Figure 4 Alt Text: The histogram of relative variability of order to pos series indicative of bullwhip effect, calculated by $\frac{\sigma^2_{order}}{\sigma^2_{pos}}$ for all promoted (Pro) and non-promoted (No Pro) series.

methodology to shed light on the comparative performance of the order and POS-based forecasts and in Section 5, we provide empirical results and show how POS and order characteristics may contribute to the performance of the forecasting models.

4 Methods and experiment setup

In this section, we describe the experimental setup and forecasting methods. Since some of the order and POS series are impacted by promotion, we use methods with the ability to incorporate promotional information. There are various statistical and machine learning methods in the literature that can be used to model the impact of causal variables. While each method has its pros and cons, there is no unique model that can operate well on all types of data and all sorts of problems (Abolghasemi, Hurley, Eshragh & Fahimnia 2020). We choose R-ARIMA and ARIMAX models as two well-known models that have shown promising results in forecasting sales time series, and in particular demand and order time series that are impacted by promotions (Abolghasemi, Beh, Tarr & Gerlach 2020). Both methods can encapsulate the explanatory variable and incorporate time series dynamic to capture time series variations. We also implement ARIMA and ETS as two basic benchmark models to ensure ARIMAX and R-ARIMA are appropriate forecasting models and outperform simple benchmark models. Both methods use price as the explanatory variable. The best forecasting models are chosen according to Akaike Information Criterion (AIC) where AIC is minimised. These methods are discussed in Sections 4.1 and 4.2, respectively.

4.1 Regression with ARIMA errors

R-ARIMA is essentially a regression model that fits an ARIMA(p,d,q) model to the residuals to capture any possible undefined correlation in residuals. The parameters p, d, and q represent the order of Autoregressive (AR) component, differencing, and Moving Average (MA) components, respectively. We use two separate R-ARIMA methods, one uses historical orders of the DCs and the other one uses the aggregated-POS data at the DCs' level. The forecasting model regresses the

orders or POS against the price of products and fits an ARIMA process to the errors as shown in Equation 1.

$$F_t = \beta_0 + \beta_1 r_t + \nu_t, \tag{1}$$

where F_t denotes forecast at time t. The parameter ν_t is an ARIMA process, and r_t is the price at time t.

Note that price carries the promotional information for each product and it is available for the next eight weeks in advance.

These methods are implemented in R and the '*forecast*' package is used to optimize and fit the parameters of methods (Hyndman, Athanasopoulos, Bergmeir, Caceres, Chhay, O'Hara-Wild, Petropoulos, Razbash, Wang & Yasmeen 2018). We choose the best model with minimum AIC and test residuals to ensure they are uncorrelated.

4.2 ARIMAX

We can construct an ARIMAX model by adding an explanatory variable to the well-known ARIMA(p, d, q) model. Price of the products inherently carries promotional information and it is known in advance. We use price as the explanatory variable in the ARIMAX model shown in Equation 2.

$$Y_t = \beta_1 r_t + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + e_t + \theta_1 e_{t-1} + ... + \theta_q e_{t-q}, \tag{2}$$

where Y_t is forecast at time t, y_{t-1} denotes the actual values of input series at time $t-1$, r_t is price at time t, and β_1 is price coefficient. The parameters ϕ and θ are the coefficients of AR and MA components, and the parameters p and q represent the orders of AR and MA components, respectively.

We use the '*TSA*' package in R to train and estimate the parameters of ARIMAX models (Chan & Ripley 2018). Overall, the order of p and q for our fitted models vary between 1 and 4. We choose the best model by minimising the AIC and test residuals to ensure they are uncorrelated.

17

4.3 Benchmark methods

We use three benchmark methods in order to illustrate the forecasting performance of the proposed methods. We first compare the forecast accuracy of R-ARIMA and ARIMAX with those of ARIMA and ETS, since these are widely accepted as standard benchmark methods in academia and practice (Hyndman & Athanasopoulos 2014). Both methods are implemented in R and the '*forecast*' package is used to fit and estimate their parameters (Hyndman, Athanasopoulos, Bergmeir, Caceres, Chhay, O'Hara-Wild, Petropoulos, Razbash, Wang & Yasmeen 2018). We state that ARIMA and ETS models are only used for benchmarking purposes. They may not be appropriate models to capture promotional periods effectively. Additionally, we benchmark our forecasts against those generated by the manufacturer.

4.4 Forecasting accuracy

We evaluate the forecasting accuracy using the relative mean absolute error (RelMAE) which is a scale-independent accuracy metric (Davydenko & Fildes 2013). Essentially RelMAE is the ratio of the error of the forecasting methods to the baseline method. We calculate RelMAE using Equation 3.

$$RelMAE_i = \frac{\sum_{t=n+1}^{n+h} |\hat{Y}_{it} - Y_t|}{\sum_{t=n+1}^{n+h} |\hat{Y}_{bt} - Y_t|}, \tag{3}$$

where h is the forecasting horizon, n is the sample size, Y_t is the observed value of the series at time t, $\hat{Y}_{it}$ is the forecast of method i at time t, and $\hat{Y}_{bt}$ is the forecast of baseline method at time t. We use ARIMA model (fitted with orders data) as the baseline statistical model.

The first 104 weeks are used as the training set and the last eight weeks are used as the test set to compare the forecasting accuracy of the models. We consider eight weeks ahead forecasts as it is of interest for the company and it provides enough time for manufacturing and logistics plannings. Moreover, price is also available only for the next eight weeks, so we can effectively forecast only for the next eight weeks. There are 484 of the products that are impacted by promotions and 200 of the products that are not impacted by any promotions. There are 71,608 observations in total from

which 9,756 observations (12% of the periods) are promotional periods. In total, we evaluate 5,472 observations in test set from which 3,923 observations are non-promotional and 1,549 observations are promotional periods. The test set corresponds to September and October 2018 and contains data with various features such as trend, seasonality, promotion status, and calendar events; representing a comprehensive dataset for evaluation. We use a pairwise t-test to determine whether there is any significant difference between POS- and order-based forecasts at 5% of the significance level.

4.5   Testing the statistical significance of time series characteristics

In order to investigate the impact of POS and order series characteristics on the accuracy of POS- and order-based methods, we use an HLM. The HLM suits our data well since demands are nested within each DC and they may have a similar ordering process (Hartzel & Wood 2017). The HLM model replaces the ordinary least squares model which is not suitable in this case since it assumes orders are independent. The HLM takes care of the possible dependency between observations at each DC by considering fixed and random effects on the regression model. The fixed effect examines the direct impact of the explanatory variable and random effect accounts for any possible group effects. The HLM model is implemented in R and the parameters are estimated using the *lme4* package (Bates et al. 2015).

5   Results and discussion

We present our empirical results in two sections. In Section 5.1, we evaluate the performance of forecasts using either POS or order series and summarise the results. In Section 5.2, we analyze ten characteristics of the POS and order series to identify whether and how they contribute to forecast performance.

5.1   POS-based vs. order-based forecasts

Table 2 shows the obtained results computed based on the mean of RelMAE for each of the discussed forecasting models. The results are summarised across promoted, non-promoted, and all SKULs. We also statistically test whether order-based forecasts have significantly different accuracy from the POS-based ones. The differences in percentages show the superiority of orders to POS-based forecasts. Note that, the company's model uses both order and POS data as described in Section 3, thus the accuracy of the forecasts are shown underneath both columns.

Table 2: Forecasts accuracy: mean of RelMAE (* significant at 5% level).

Forecasting methods	Promoted			Non Promoted			All SKULs		
	Order	POS	Diff.	Order	POS	Diff.	Order	POS	Diff.
ARIMAX	0.87	0.98	11%*	0.88	1.04	15%*	0.88	1.03	14%*
R-ARIMA	0.88	1.01	12%*	0.89	1.04	14%*	0.89	1.04	14%*
ARIMA	1.00	1.03	3%	1.00	1.10	10%*	1.00	1.04	4%
ETS	1.01	1.04	4%	1.01	1.07	6%*	1.02	1.06	3%
Company's model	0.96	0.96	-	0.99	0.99	-	0.99	0.99	-

Our results show that the order-based forecasts, on average, outperform the POS-based forecasts in forecasting the DCs' orders by between 6 to 15%. We observe that the superiority of order-based models decreases in the presence of promotions. For the promoted SKULs, the results show that the order-based forecasts significantly generate more accurate forecasts (at 5% significance level) than the POS-based forecasts when the ARIMAX and R-ARIMA methods are utilized to forecast the DCs' orders. However, when the ARIMA and ETS methods are employed to forecast the DCs' orders for promoted products, there is no significant difference between the accuracy of the forecasts that use either order or POS series. This is because neither of the univariate ARIMA and ETS methods incorporates promotional information to forecast the DCs' orders and these methods may not be suitable to forecast our order and POS series that are impacted by promotion. Consequently,

they both have large forecasting errors and their forecasts are not significantly different. For the non-promoted SKULs, there is a statistically significant difference at a 5% level between the forecast accuracy of all employed methods when using POS or order series.

The ARIMAX and R-ARIMA methods depict higher accuracy than the ARIMA and ETS methods. This is of no surprise since both the ARIMAX and R-ARIMA methods use price as additional information to forecast orders as opposed to ARIMA and ETS that are univariate methods and only use previous values of series. This indicates that additional information about price can be useful only if it is added and modeled properly to the model. The accuracy of the company's forecasting method, on average, is superior to the ARIMA and ETS methods. This implies that using POS information in a subjective manner (in the form of expert's judgments) adds value to the company's statistical forecasts. However, the ARIMAX and R-ARIMA methods outperform the company's method, both in the absence and presence of promotions, indicating that choosing an appropriate type of forecasting method that can effectively and objectively take into account the POS information is beneficial.

We also investigate the frequency of the periods that one model outperforms the other one. The order-based forecasts using the ARIMAX and R-ARIMA methods outperform the POS-based ones over 68% and 61% of the periods, respectively. This ratio drops to 51% and 49% for the ARIMA and ETS methods, respectively. In order to know what factors make one model more accurate than the other one for a particular series, we analyze their behavior using various characteristics as discussed in Section 3.1 and provide a thorough discussion in Section 5.2.

Our results with regards to the superiority of order-based forecasts to the POS-based ones in upstream SCs are consistent with recent findings by Narayanan et al. (2019). However, our experiment differs from their study and adds value in the following ways: i) while their finding is only based on 21 products and includes partially synthetic data, we use a large real-world dataset with 684 SKULs, ii) POS and order series used in our study contain different features, depict different behavior and includes promotions as an important feature, and iii) we employ different forecasting methods in our experiment.

We summarise our findings as follows: First, it is imperative to choose the right forecasting method to be able to compare the value of the POS and order series. We showed that if we choose inappropriate models to forecast the series, there will be no statistically significant difference between the accuracy of the forecasts regardless of the chosen input series. Second, for an appropriate choice of a forecasting model, using order series results in higher forecast accuracy than using the POS data to forecast the DCs' orders. However, this superiority declines in the presence of promotions and may depend on the behavior and characteristics of the series. Third, in our case study, we observe that using POS data in a judgmental fashion can improve the forecasting accuracy of a statistical model that is merely based on historical order series. However, POS data can be used more objectively in a quantitative model to forecast the manufacturer's demand. This can result in higher forecasting accuracy without any intervention.

5.2 Discussion on the characteristics of order and POS series

The results of the M4 forecasting competition and other studies verify that the characteristics of time series impact the forecasting accuracy of methods (Wang et al. 2009, Kang et al. 2017, Spiliotis et al. 2020). Now we turn our attention to determining the significant characteristics of the POS and order series and their corresponding impact on the accuracy of the forecasts. It is of great importance to know the characteristics of the POS and order series and circumstances under which using one source of data is superior to one another in forecasting orders. We found that there is no significant difference between the forecasting accuracy of the ARIMAX and R-ARIMA methods as two top-performing methods when the same source of data, i.e., either POS or order series, is used to generate forecasts. Therefore, we only analyze the forecasts generated by the ARIMAX method. We evaluate the performance of the POS-based and the order-based forecasts and investigate their association with POS and order characteristics using an HLM model. We measure the relative accuracy of order-based to the POS-based forecasts by $MAE(POS)/MAE(Order)$, and regress this dependent variable against the characteristics of all POS and order series as fixed independent variables. Although we can not calculate these characteristics for the promotional and

non-promotional observations individually, we take into account the promotion frequency and promotion impact by considering the ratio of promotional periods as another independent variable. Note that the promotion ratio for POS and order series is identical and it is indicated with a single variable in the HLM model. We fit one global model on all 684 series characteristics and report the results. Table 3 shows the variables, their estimated coefficient, standard error, t-value, and their p-values. We obtain the p-values with the likelihood ratio test where we tested the full model against the reduced model of the variable that we are interested in.

Our research findings demonstrate that the mean of POS series, the variance of the order and POS series, the entropy and non-linearity of order and POS series, and the frequency of promotions are the characteristics of the series that have a statistically significant impact on the comparative performance of the POS and order data in forecasting DCs' orders. In our experiment, the other investigated characteristics of series including trend, seasonality strength, hurst, stability, kurtosis, and skewness of POS and order series do not significantly impact the relative performance of the POS- to order-based forecasts.

According to the results in Table 3, the relative MAE of the POS-based forecasts to order-based forecasts decreases when the POS mean increases, i.e., the relative performance of POS-based forecasts increases as POS mean increases. We attribute this to the intermittency nature of order series when the POS series have a small mean. We scrutinized the POS and order series and found that the SKULs with very low volumes of POS correspond to the SKULs with lower levels of orders frequency. For these series, there is no order placed at some periods, i.e., order series have intermittent nature with low intermittency level. Since the DCs' orders depict intermittency, using POS data to forecast the DCs' orders may not be suitable. This is consistent with the results that are reported at the retailers level. According to (Williams & Waller 2010), the higher volume of sales improves the relative performance of the POS-based methods to the order-based methods when forecasting retailers' orders.

The variance of the POS and order series are the other statistically significant characteristics of the series that can impact the relative performance of POS- and order-based forecasts. Our results show

23

Table 3: The HLM model results (*significant at 5% level, ** significant at 10% level)

	Input	Coefficient	Std. Error	t value	P-values
(Intercept)		1.13	0.29	3.89	0.00*
Mean	POS	-0.21	0.09	-2.31	0.02*
	Orders	1.51	0.84	1.79	0.07**
Variance	POS	-0.29	1.55	-0.18	0.01*
	Orders	0.64	1.49	0.43	0.05*
Trend	POS	0.17	0.26	0.67	0.50
	Orders	-0.06	0.22	-0.30	0.76
Hurst	POS	-0.18	0.11	-1.57	0.11
	Orders	-0.03	0.09	-0.33	0.73
Stability	POS	-0.29	0.27	-1.05	0.29
	Orders	0.15	0.24	0.64	0.52
Seasonal strength	POS	0.13	0.07	1.69	0.09
	Orders	0.01	0.08	0.15	0.87
Non-linearity	POS	-0.38	0.20	-1.90	0.06**
	Orders	0.20	0.09	2.06	0.04*
Entropy	POS	-0.23	0.20	-1.15	0.06**
	Orders	0.01	0.22	0.03	0.07**
Kurtosis	POS	0.29	0.12	2.46	0.14
	Orders	-0.22	0.11	-1.91	0.55
Skewness	POS	0.01	0.16	0.08	0.93
	Orders	0.03	0.15	0.21	0.82
Promotion	POS	-0.05	0.02	-2.21	0.03*

that the relative MAE of the POS-based to order-based forecasts is negatively correlated with the variance of the POS series. That is, the relative performance of the POS-based to order-based forecasts increases when the variability of the POS series increases. This may attribute to the promotion impact and the order smoothing during promotions that cause anti-bullwhip behavior in series. When POS series are impacted by large promotions they exhibit large variability. The corresponding order series depict smoother behavior and smaller variance as DCs smooth their orders during promotional periods. However, DCs change their orders frequently to match their actual demand. As such, the orders placed by the DCs may be disturbed and they may not be a good signal for forecasting DCs' orders, instead, POS information can be more helpful to forecast the uncertain orders. The same results hold for non-linearity, i.e., the relative MAE of POS-based to order-based forecasts decreases when the non-linearity of the POS series increases, i.e., the performance of POS-based forecasts increases as POS variability increases. Non-linearity in POS series is mainly caused by promotion and our results show that when series depict higher values of non-linearity, it is better to use POS series that directly carry promotional information rather than using DCs' orders that are disturbed by DCs' forecasts error and operational constraints.

The entropy of POS and order series was also found to be another characteristic of the POS and order series that significantly impact their forecasting accuracy. Our results show that the relative MAE of the POS-based to order-based forecasts is negatively correlated with the entropy in the POS series. That is, the POS-based forecasts tend to perform better when the entropy of POS series increases. This may attribute to the DCs' smoothing during promotions and in general DCs' forecasting accuracy. Since noisy data are less forecastable, DCs may have less accurate forecasts when POS data are noisy and accordingly change their orders to match with their actual demand. DCs and retailers are not usually as good as their suppliers in forecasting (Gilliland 2010). As such, the orders placed by the DCs may be disturbed and they may not be a good signal for forecasting. Thus, it might be better to rely on POS series for forecasting.

According to our results, the negative coefficient of the promotion factor indicates that the relative MAE of the POS-based forecasts to order-based forecasts decreases in the presence of promotions,

i.e., the performance of POS-based forecasts increases in the presence of promotion. One explanation is that orders and POS series are more similar in the absence of promotion. However, DCs smooth their orders during promotions and POS data at each period does not directly translate to the orders but span over a few periods with uncertain values that depend on operational factors and the ordering policy of the DCs. This may render POS data as an appropriate candidate to forecast the DCs' orders during promotions. This finding aligns with the current practice in the company and experts' opinions who state that POS information is a valuable source of information to forecast DCs' orders during promotions. This finding may be limited to our study and attributed to the promotion and the specific ordering policy of DCs that occurs during promotions. The promotions impact on the value of the POS and order series at different stages of the SCs need to be further investigated.

In summary, we show that using the order series to forecast manufacturer's demand, on average, results in higher forecast accuracy than using the POS series. However, their relative performance is tied to the characteristics of the input series and may change from one series to another. We found that the mean, variance, non-linearity and entropy of POS series, and promotion status are statistically significant features that can impact the value of the POS data for forecasting orders. We state that the usefulness of the POS or order series for forecasting may change across the SCs and managers should not evaluate their value merely based on their nature, i.e., being POS or order. Our results show that POS data may be more adequate for order forecasting in the presence of promotion, otherwise historical orders are more appropriate for forecasting non-promotional orders. Also, one should consider various features of POS data such as mean, variance, non-linearity, and entropy to determine whether they will be useful for order forecasting or not. Table 4 shows the summary of results. As shown in Table 4, the relative performance of the POS-based to order-based forecasts tend to improve under five conditions: i) the mean of POS series increases while orders are not intermittent, ii) the variance of POS series increases, iii) the entropy of POS series increases, iv) the non-linearity of POS series increases, v) promotions are offered to customers.

Table 4: Summary of findings with regards to relative performance of POS to order based forecasts

Feature	Status	POS/Order
POS Mean	Increasing	Increasing
POS Variance	Increasing	Increasing
POS entropy	Increasing	Increasing
POS non-linearity	Increasing	Increasing
Promotion	On	Increasing

6   Conclusions

In this paper, we empirically analyze a three-echelon SC from a major FMCG food company in Australia. We focus on the upstream SC where the manufacturer aims to forecast the DCs' orders. DCs do not share their operational information such as ordering policy and inventory level with the manufacturer but the retailers POS data is available in the aggregated form at DCs level. The manufacturer can use either historical orders or aggregated POS data to forecast the DCs' orders. The order and POS series are highly impacted by promotions and depict different behaviors.

In order to investigate the value of POS information sharing in upstream SC forecasting, we compare the accuracy of POS- and order-based methods in forecasting the DCs' orders. We empirically show that the order-based forecasts outperform the POS-based forecasts by between 6%-15% (on average) in forecasting the DCs' orders. We illustrate that the comparative value of the POS and order data depends on the type of forecasting method and series characteristics.

We use four different forecasting methods to forecast the DCs' orders. Since POS and order series are impacted by promotion, we use the ARIMAX and R-ARIMA methods to forecast the DCs' orders and compare them with the common benchmark methods, ARIMA and ETS. We empirically analyze their performance using the orders and POS series for 684 SKULs and show that our proposed methods can effectively take into account the POS information to forecast the DCs' orders. We also compare our model with the employed model in the case study and show that our proposed

27

ARIMAX and R-ARIMA methods outperform the company's forecasts that incorporate the POS data subjectively into the output of its order forecasting model.

We further analyze ten important characteristics of the orders and the POS time series to determine how information evolves across the SCs. We show that the distribution of orders and POS are similar in various characteristics but they depict different levels of strength. We determine the significant characteristics of the POS and order series that contribute to their forecasting performance. We illustrate that order-based models are, on average, more accurate than the POS-based ones in upstream SCs. However, the performance of the POS-based forecasts tend to improve during promotions and when POS series depict higher mean, variance, entropy, and non-linearity.

The results of this research shed light on the value of the POS data in upstream SCs forecasting and provide guidelines for practitioners on using the suitable types of input series for forecasting DCs' orders. There are some limitations to this study that may be considered for future research. Primarily, our insights are tied to the current dataset that is mostly cereal and breakfast products with an FMCG nature. Although the models and results are analyzed on a large dataset to ensure its robustness and validity for a dataset with similar characteristics, the results may not be applicable for other datasets such as slow-moving products. Thus, replicating the experiment on other datasets with different characteristics may reveal new insights. In addition, in this study we focused on a weekly dataset but a natural extension to this study is to consider temporally aggregated data, e.g., monthly, and investigate whether the same results hold or not. We also suggest that theoretical analysis can be beneficial to build theories in this area. The other important research avenue is to further evaluate the impact of the promotions. While promotions have been extensively studied in the forecasting literature, it is less considered in the information sharing literature and needs to be further investigated both in theoretical and empirical research. The utility of forecasts in SCs is another important research direction. The higher accuracy of forecasts can impact SCs performance in different ways and eventually result in a lower inventory level, higher service level, and lower cost. However, the improvement in forecasting accuracy does not necessarily reduce the inventory costs and the relationship between these two need to be investigated more (Syntetos & Boylan 2006).

Finally, we use either the POS or order series individually to forecast DCs' orders. Williams et al. (2014) used both POS and order data to predict retailer's orders and showed that their approach improves the forecasting accuracy by 125 %. They found that the retail echelon inventory processes translate into a long-run equilibrium between order and POS called the inventory balance effect. However, their approach does not accommodate promotions impact and is not suitable if a cointegration does not exist for POS and order series. The presence of promotion and order smoothing changes the dynamic of POS and order series, resulting in conditions where cointegration of them may not be stationary. Van Belle, Guns & Verbeke (2020) investigated 50 products from the pharmacological industry and found that using both sell-through (POS) and order series can improve the forecast accuracy at the manufacturer's level by 4.7%, on average, and at most by 10%. Different results are reported in the literature and more studies are required to examine the value of using both POS and order series simultaneously to forecast DCs' orders.

Acknowledgment

Data Statement Availability

The data that support the findings of this study are available on request from the corresponding author subject to permission from data owners. The data are not publicly available due to containing confidential information that could compromise the privacy of stakeholders.

References

Abolghasemi, M., Beh, E., Tarr, G. & Gerlach, R. (2020), 'Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion', *Computers & Industrial Engineering* p. 106380.

Abolghasemi, M., Hurley, J., Eshragh, A. & Fahimnia, B. (2020), 'Demand forecasting in the presence of systematic events: Cases in capturing sales promotions', *International Journal of Production Economics* **230**, 107892.

Abolghasemi, M., Hyndman, R. J., Spiliotis, E. & Bergmeir, C. (2022), 'Model selection in reconciling hierarchical time series', *Machine Learning* pp. 1–51.

Alftan, A., Kaipia, R., Loikkanen, L. & Spens, K. (2015), 'Centralised grocery supply chain planning: improved exception management', *International Journal of Physical Distribution & Logistics Management* **45**(3), 237–259.

Aviv, Y. (2001), 'The effect of collaborative forecasting on supply chain performance', *Management Science* **47**(10), 1326–1343.

Barratt, M. (2003), 'Positioning the role of collaborative planning in grocery supply chains', *The International Journal of Logistics Management* **14**(2), 53–66.

Bassamboo, A., Moreno, A. & Stamatopoulos, I. (2017), 'Inventory auditing and replenishment using point-of-sales data', *Production and Operations Management* .

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015), 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software* **67**(1), 1–48.

Chan, H. K. & Chan, F. T. (2009), 'Effect of information sharing in supply chains with flexibility', *International Journal of Production Research* **47**(1), 213–232.

Chan, K.-S. & Ripley, B. (2018), *TSA: Time Series Analysis.* R package version 1.2.

URL: *https://CRAN.R-project.org/package=TSA*

Choudhary, D. & Shankar, R. (2015), 'The value of vmi beyond information sharing under time-varying stochastic demand', *International Journal of Production Research* **53**(5), 1472–1486.

Christopher, M. & Holweg, M. (2011), 'Supply chain 2.0 : Managing supply chains in the era of turbulence', *International Journal of Physical Distribution & Logistics Management* **41**(1), 63–82.

Claassen, M. J., Van Weele, A. J. & Van Raaij, E. M. (2008), 'Performance outcomes and success factors of vendor managed inventory (VMI)', *Supply Chain Management: An International Journal* **13**(6), 406–414.

Cover, T. M. (1999), *Elements of information theory*, John Wiley & Sons.

Croson, R. & Donohue, K. (2003), 'Impact of POS data sharing on supply chain management: An experimental study', *Production and Operations Management* **12**(1), 1–11.

Davydenko, A. & Fildes, R. (2013), 'Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts', *International Journal of Forecasting* **29**(3), 510–522.

Disney, S. M. & Towill, D. R. (2003), 'Vendor-managed inventory and bullwhip reduction in a two-level supply chain', *International journal of operations & production Management* **23**(6), 625–651.

Fulcher, B. D. & Jones, N. S. (2017), 'hctsa: A computational framework for automated time-series phenotyping using massive feature extraction', *Cell systems* **5**(5), 527–531.

Gilliland, M. (2010), *The business forecasting deal: exposing myths, eliminating bad practices, providing practical solutions*, Vol. 27, John Wiley & Sons.

Goerg, G. (2013), Forecastable component analysis, *in* 'International Conference on Machine Learning', pp. 64–72.

Hartzel, K. S. & Wood, C. A. (2017), 'Factors that affect the improvement of demand forecast accuracy through point-of-sale reporting', *European Journal of Operational Research* **260**(1), 171–182.

Hosoda, T., Naim, M. M., Disney, S. M. & Potter, A. (2008), 'Is there a benefit to sharing market sales information? Linking theory and practice', *Computers & Industrial Engineering* **54**(2), 315–326.

Hyndman, R. J. & Athanasopoulos, G. (2014), *Forecasting: principles and practice*, OTexts.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. & Yasmeen, F. (2018), 'forecast: Forecasting functions for time series and linear models'.

Hyndman, R., Kang, Y., Montero-Manso, P., Talagala, T., Wang, E. & Yang, Y. (2018), *tsfeatures: Time Series Feature Extraction*. R package version 1.0.0.
URL: *https://pkg.robjhyndman.com/tsfeatures/*

Kang, Y., Hyndman, R. J. & Smith-Miles, K. (2017), 'Visualising forecasting algorithm performance using time series instance spaces', *International Journal of Forecasting* **33**(2), 345–358.

Lau, J. S., Huang, G. Q. & Mak, K. (2004), 'Impact of information sharing on inventory replenishment in divergent supply chains', *International Journal of Production Research* **42**(5), 919–941.

Mishra, B. K., Raghunathan, S. & Yue, X. (2007), 'Information sharing in supply chains: Incentives for information distortion', *IIE Transactions* **39**(9), 863–877.

Narayanan, A., Sahin, F. & Robinson, E. P. (2019), 'Demand and order-fulfillment planning: The impact of point-of-sale data, retailer orders and distribution center orders on forecast accuracy', *Journal of Operations Management* **65**(5), 468–486.

Perry, M. & Sohal, A. S. (2001), 'Effective quick response practices in a supply chain partnership-An Australian case study', *International Journal of Operations & Production Management* **21**(5/6), 840–854.

Pollock, D. S. G. (2001), 'Methodology for trend estimation', *Economic Modelling* **18**(1), 75–96.

Rose, O. (1996), 'Estimation of the hurst parameter of long-range dependent time series', *University of Wurzburg, Institute of Computer Science Research Report Series* .

Seifert, D. (2003), *Collaborative planning, forecasting, and replenishment: How to create a supply chain advantage*, AMACOM Div American Mgmt Assn.

Singh, H., Garg, R. & Sachdeva, A. (2018), 'Supply chain collaboration: A state-of-the-art literature review', *Uncertain Supply Chain Management* **6**(2), 149–180.

Spiliotis, E., Kouloumos, A., Assimakopoulos, V. & Makridakis, S. (2020), 'Are forecasting competitions data representative of the reality?', *International Journal of Forecasting* **36**(1), 37–53.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S. & Nikolopoulos, K. (2016), 'Supply chain forecasting: Theory, practice, their gap and the future', *European Journal of Operational Research* **252**(1), 1–26.

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

Teräsvirta, T., Lin, C.-F. & Granger, C. W. (1993), 'Power of the neural network linearity test', *Journal of Time Series Analysis* **14**(2), 209–220.

Van Belle, J., Guns, T. & Verbeke, W. (2020), 'Using shared sell-through data to forecast wholesaler demand in multi-echelon supply chains', *European Journal of Operational Research* .

Wang, N., Lu, J., Feng, G., Ma, Y. & Liang, H. (2016), 'The bullwhip effect on inventory under different information sharing settings based on price-sensitive demand', *International Journal of Production Research* **54**(13), 4043–4064.

Wang, X., Smith, K. & Hyndman, R. (2006), 'Characteristic-based clustering for time series data', *Data Mining and Knowledge Discovery* **13**(3), 335–364.

Wang, X., Smith-Miles, K. & Hyndman, R. (2009), 'Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series', *Neurocomputing* **72**(10-12), 2581–2594.

Williams, B. D. & Waller, M. A. (2010), 'Creating order forecasts: point-of-sale or order history?', *Journal of Business Logistics* **31**(2), 231–251.

Williams, B. D. & Waller, M. A. (2011), 'Top-down versus bottom-up demand forecasts: the value of shared point-of-sale data in the retail supply chain', *Journal of Business Logistics* **32**(1), 17–26.

Williams, B. D., Waller, M. A., Ahire, S. & Ferrier, G. D. (2014), 'Predicting retailer orders with POS and order data: The inventory balance effect', *European Journal of Operational Research* **232**(3), 593–600.

A Time series Characteristics

1. Mean: Mean is the arithmetic average of a time series.

2. Variance: Variance is the spread of observations in a time series. We calculate the variance of orders and POS for each SKUL. The ratio of orders variance (σ^2_{order}) to POS variance (σ^2_{pos}) indicate the bullwhip effect index for a particular SKUL (Disney & Towill 2003).

3. Trend: Trend refers to a process where time series mean changes over long term periods. Consider a time series $Y_t = S_t + T_t + E_t$, where S_t, T_t, and E_t denote the seasonal, trend, and errors, respectively. We can detrend and deseasonalise a time series by subtracting the trend, and seasonality components from series. Let $X_t = Y_t - T_t$ and $Z_t = X_t - S_t$ represent the detrended and deeseasonalised components, respectively. Then, the remainder is error term, $e = Y_t - T_t - S_t$. Trend strength can be measured by Equation 4 (Pollock 2001, Wang et al. 2009).

$$\text{Trend strength} = 1 - \frac{\text{Var}(e_t)}{\text{Var}(Z_t)} \tag{4}$$

4. Hurst: Hurst measures the long term dependence of a time series. This measure can be estimated with different methods (Rose 1996). In order to compute hurst, an autoregressive fractionally integrated moving average model is fitted to the time series where d is the degree of first differencing (Hyndman, Kang, Montero-Manso, Talagala, Wang & Yang 2018). Hurst is equal to 0.5 plus the maximum likelihood estimate of the fractional differencing order d to make it consistent with hurst coefficients (Wang et al. 2009).

5. Stability: Stability measures the variance of means of series. We first calculate the mean of POS and order for each series, then calculate the variance of the means across all series. This metric gives statistics about the spread of means of POS and order series.

6. Seasonality strength: Time series depict seasonality patterns when they exhibit a pattern that is repetitive and is often caused by seasonal factors. Time series with a fixed seasonality has a

large autocorrelation at fixed seasonal lags. Seasonality strength can be measured by Equation 5 (Wang et al. 2009).

$$\text{Seasonality strength} = 1 - \frac{\text{Var}(e_t)}{\text{Var}(X_t)} \tag{5}$$

7. Non-linearity: This metric measures the strength of the non-linearity relationship between observations of a time series. We measured non-linearity with the modified version of the Teräsvirta's non-linearity test as described in (Hyndman, Kang, Montero-Manso, Talagala, Wang & Yang 2018, Teräsvirta et al. 1993).

8. Entropy: This measures the "forecastability" of a time series. Large values of entropy indicate a high noise-to-signal ratio that makes them more difficult to forecast (Goerg 2013). Entropy is the *Shanon entropy* as shown in the Equation (6),

$$\text{entropy} = -\int_{-\pi}^{\pi} \hat{f}(\lambda) \log(\hat{f}(\lambda)) d\lambda, \tag{6}$$

where $\hat{f}(\lambda)$ is the spectral density of the data. The spectral density describes the power or strength of a time series as a function of frequency λ.

9. Kurtosis: Kurtosis measured whether the distribution of data is heavy-tailed or light-tailed. Higher values of kurtosis represent time series with heavy tails and lower levels of kurtosis indicate time series with light tails. We measure kurtosis with Equation 7 as shown below:

$$\text{Kurtosis} = \frac{1}{n\sigma^4} \sum_{t=1}^{n} (Y_t - \bar{Y})^4, \tag{7}$$

where n is the number of observations, and $\bar{Y}$ and σ represent the average and standard deviation of the time series, respectively.

10. Skewness: Essentially skewness measures the extent by which a distribution is deviated from a normal distribution. We measured the skewness using Equation 8:

$$\text{Skewness} = \frac{1}{n\sigma^3} \sum_{t=1}^{n} (Y_t - \bar{Y})^3, \tag{8}$$

where n is the number of observations, and $\bar{Y}$ and σ represent the average and standard deviation of the time series, respectively.

For more information about these characteristics refer to (Wang et al. 2006, Kang et al. 2017).