This article was downloaded by: [155.246.103.35] On: 06 April 2017, At: 00:30
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Information Sharing in Supply Chains: An Empirical and Theoretical Valuation

Ruomeng Cui, Gad Allon, Achal Bassamboo, Jan A. Van Mieghem

To cite this article:
Ruomeng Cui, Gad Allon, Achal Bassamboo, Jan A. Van Mieghem (2015) Information Sharing in Supply Chains: An Empirical and Theoretical Valuation. Management Science 61(11):2803-2824. http://dx.doi.org/10.1287/mnsc.2014.2132

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2015, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

MANAGEMENT SCIENCE
Vol. 61, No. 11, November 2015, pp. 2803–2824
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.2132
© 2015 INFORMS

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Information Sharing in Supply Chains: An Empirical and Theoretical Valuation

Ruomeng Cui
Kelley School of Business, Indiana University, Bloomington, Indiana 47405,
cuir@indiana.edu

Gad Allon, Achal Bassamboo, Jan A. Van Mieghem
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208,
{g-allon@kellogg.northwestern.edu, a-bassamboo@kellogg.northwestern.edu, vanmieghem@kellogg.northwestern.edu}

We provide an empirical and theoretical assessment of the value of information sharing in a two-stage supply chain. The value of downstream sales information to the upstream firm stems from improving upstream order fulfillment forecast accuracy. Such an improvement can lead to lower safety stock and better service. Based on the data collected from a consumer packaged goods company, we empirically show that, if the company includes the downstream sales data to forecast orders, the improvement in the mean squared forecast error ranges from 7.1% to 81.1% across all studied products. Theoretical models in the literature, however, suggest that the value of information sharing should be zero for over half of our studied products. To reconcile the gap between the literature and the empirical observations, we develop a new theoretical model. Whereas the literature assumes that the decision maker strictly adheres to a given inventory policy, our model allows him to deviate, accounting for private information held by the decision maker, yet unobservable to the econometrician. This turns out to reconcile our empirical findings with the literature. These "decision deviations" lead to information losses in the order process, resulting in a strictly positive value of downstream information sharing. Furthermore, we empirically quantify and show the significance of the value of operations knowledge—the value of knowing the downstream replenishment policy.

Keywords: supply chain; information sharing; signal propagation; decision deviation; time series; empirical forecasting; autoregressive integrated moving average process
History: Received April 9, 2013; accepted November 25, 2014, by Serguei Netessine, operations management. Published online in *Articles in Advance* June 29, 2015.

1. Introduction

The abundance of information technology has had a massive impact on supply chain coordination. Sharing downstream demand information with upstream suppliers has improved supply chain performance in practice. Costco and 7-Eleven share warehouse-specific, daily, item-level, point-of-sale data with their suppliers via SymphonyIRI platform, a company offering business advice to retailers (see *Retail Info Systems News* 2013). In addition to this unidirectional information sharing, collaborative planning, forecasting, and replenishment (CPFR) programs advocate joint visibility and joint replenishment. According to Terwiesch et al. (2005), the benefit of CPFR programs can be significant: the GlobalNetXchange, a consortium consisting of more than 30 trade partners, has reported a 5%–20% reduction in inventory costs and an increase in off-the-shelf availability of 2%–12% following the launch of their CPFR programs.

Companies spend billions of dollars on demand forecasting software and other supply chain solutions (Ledesma 2004). Given the implementation cost of collaboration technology and the limited theoretical benefits, it is not clear in practice whether a firm should invest in information sharing systems. The decision to implement an information sharing system thus hinges on the following question: How much would sharing downstream sales information improve the supplier's order forecast accuracy? We were approached with this question by the statistical forecasting team of a leading global consumer packaged goods (CPG) company that manufactures and sells beverages and snack foods to wholesalers and retail chains. Forecasting is necessary for the company because of the lead time to adjust manufacturing runs and deploy inventory. In the absence of downstream sales information, the upstream supplier uses its own sales history (i.e., its retailer's order history), to forecast how much to manufacture. Not satisfied with its current forecasting performance, the firm sought solutions in information sharing by collecting downstream operations data (e.g., point of sale) from its customers. Using this

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

data set, we directly measure the supplier's forecast accuracy improvement.

Our empirical results indicate a substantial value of information sharing: statistically significant improvements (7.1% to 81.1% mean squared error percentage improvements) across almost all studied products. To put this in perspective, the company views forecast accuracy improvement opportunities of 10% as important and 30% as very significant.

The benefits of information sharing have been theoretically quantified in the literature. The works of Gaur et al. (2005) and Giloni et al. (2014) are important antecedents of our paper. The authors find that the value of sharing downstream sales to improve upstream forecasting is limited. (We will also refer to customer sales as *customer demand* or *demand*.) In their setting, the decision maker strictly follows an order-up-to policy, via which the demand process propagates upstream and becomes the order process. If, for example, the retailer follows a demand replacement policy (the retailer orders the demand in the current week), orders equal demand. It is as if demand propagates fully upstream and orders carry full demand information. In such settings, there is no value of information sharing. The insights in the literature show that the value of information sharing is zero when the upstream order is a sufficient statistic of demand.

Although our empirical observations show that there is value of information sharing, the theory (which follows the same spirit as Gaur et al. 2005 and Giloni et al. 2014) would suggest zero value of information sharing for 10 out of 14 studied products. These different results suggest that we need a better theoretical understanding of the missing component in the theoretical literature, which makes the results in the literature no longer apply.

The key underlying assumption in the theoretical literature is that decision makers consistently and strictly follow a given replenishment policy. In practice, however, we learned that decision makers deviate from their target inventory policy based on private information that we cannot observe. From an econometric perspective, we model the agent's deviation from the exact policy, in the spirit of Rust (1994), by an "error term" that accounts for a state variable that is observed by the agent but not by the statistician. Taking into account the potential decision deviations from classical ordering policies significantly increases the theoretical value of information sharing, in agreement with the empirical findings.

In the presence of decision deviations, we prove that the value of information sharing is strictly positive for any forecast lead time (regardless of the demand structure and the ordering policy). At first glance, the decision uncertainty seems to diminish the attractiveness of analyzing a retailer's replenishment process because of the unpredictability of the order decision. Such uncertainty, however, opens the door to information loss as signals propagate upstream. As demand signals and decision deviations propagate upstream to produce the order process, they follow distinct evolution patterns: The evolution of inventory governs the translation of decision deviations into replenishment decisions, and the evolution of inventory and current demand together govern the translation of demand. This difference prevents orders from carrying both full information of demand and decision deviations. Information sharing then becomes valuable to recover the order's elaborate information structure and to improve forecast accuracy. This intuition continues to remain for any linear demand and order structure, and thus, our conclusion is robust under more general settings. Our new theory is supported by the empirical observations of significant forecast improvements.

We conduct comparative statics and numerical studies to examine the impact of product demand characteristics, such as the degree of seasonality, on the value of information sharing. These insights can help managers rank the potential gains from information sharing depending on the demand characteristics for different products such as sport drinks or orange juice.

Our estimation procedure uses the fact that the supplier not only knows the retailer's point-of-sale data, but also knows the retailer's replenishment policy. We refer to knowing the replenishment policy as "operational knowledge" (in contrast to "sales knowledge"). We are able to disentangle the value of sales and the value of operational knowledge. By using an analogous estimation procedure that only uses sales data (and no operational knowledge), we empirically quantify the value of sales and operational knowledge. In fact we show that operational knowledge brings the same order of magnitude of forecast improvements as only sales data. This suggests that companies should always keep operations in mind to achieve the maximum value from downstream sales information.

Our study is grounded in both empirical evidence and theory, and attempts to understand the cause of the positive value of information sharing. We analyze a data set containing weekly downstream demand, upstream order fulfillment, and the point-of-sale price over a period of two and a half years. This allows us to make the following three main contributions. First, this paper complements the emerging area of research in information sharing with empirical evidence. Specifically, we directly measure the value of information sharing at a leading CPG company and demonstrate a statistically significantly positive value of information sharing, and we empirically quantify

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

the value of operational knowledge. Second, we allow for decision deviations in our theoretical model to explicitly capture the decision maker's private information that is unobservable to us. This model extends the existing literature and recovers the results from the literature as a special case without decision deviation. We demonstrate that the decision deviation distorts the normal demand propagation in a way that obscures the detailed information of the two processes. The resulting less informative order signals induce larger forecast uncertainty, which suggests that it is strictly beneficial for the supplier to use downstream demand to recover the order's original elaborate information structure. Third, we provide guidelines on how the value of information sharing depends on the demand characteristics.

2. Literature Review

Our paper is related to two streams of literature: (1) theoretical work on information sharing and demand propagation through supply chains and (2) empirical work that bridges the above theory and operational data.

There is a vast theoretical literature on the subject of demand propagation and information sharing in supply chains. A company's demand propagates through the supply chain and becomes its order to the supplier. The properties of orders can help answer important questions in supply chains, e.g., is sharing retailer's demand information beneficial for the supplier to forecast its own order and manage its inventory? Is there incentive for the agents to share their own information? Is there a bullwhip effect and what is the driver? The demand propagation relies on two basic characteristics of the supply chain: demand structure and replenishment policy. We focus on the work that assumes truthful and complete information disclosure. We begin by introducing the various demand and policy structures studied in the literature. Next, we discuss our paper's contribution relative to the two most related studies: Gaur et al. (2005) and Giloni et al. (2014).

2.1. Theoretical Work

The demand propagation and information sharing have been well studied and quantified under various modeling assumptions. Lee et al. (2000) adopt an autoregressive AR(p) process, Miyaoka and Hausman (2004) and Graves (1999) assume an integrated moving average IMA(d, q) process, Gaur et al. (2005) and Giloni et al. (2014) consider an autoregressive moving average ARMA(p, q) process, and Aviv (2003) uses the linear state space framework. Another body of literature applies the martingale model of forecast evolution (MMFE) structure. It uses the incremental signal, generated from the minimum mean squared error, to model the evolution of a process. Heath and Jackson (1994), Graves et al. (1998), Aviv (2001), and Chen and Lee (2009) apply such demand structure to study production and forecasting. Mixed results have been derived based on different demand structures. For example, Lee et al. (2000) find the value of demand can be quite high with an autoregressive (AR)(1) demand; Gaur et al. (2005) find that there is no value of information sharing under 75% of demand parameters when demand follows an autoregressive moving average ARMA(1, 1) process.

Gaur et al. (2005) show that the ARMA model closely resembles the real-life demand structure and find it valuable from the manager perspective to study such demand process. Our study models and empirically fits an autoregressive integrated moving average (ARIMA) demand, because it is the most general structure to describe our data set.

In the information sharing literature, the most commonly studied replenishment policy is the myopic order-up-to policy. The following papers investigate other ordering policies. Caplin (1985) studies a periodically reviewed (s, S) policy and proves the existence of the bullwhip effect. Cachon and Fisher (2000) quantify the value of information sharing with a batching allocation rule between one supplier and multiple retailers. These two papers model batching in replenishment, which is not amenable to mathematically tractable analysis. The following papers adopt a "linear replenishment rule," in which orders are linear in historical observations. Balakrishnan et al. (2004) propose an "order smoothing" inventory policy where the order is a convex combination of historical demands. Miyaoka and Hausman (2004) use the old demand forecasts to set the base stock level and show this can reduce the bullwhip effect. Graves et al. (1998) and Chen and Lee (2009) study the general-order-up-to policy (GOUTP), which smooths forecast revisions to produce a desirable order-up-to level. According to the replenishment policy that the retailer adopts, our paper introduces a linear order rule that keeps the days of inventory constant and uses some order smoothing. This policy equals the optimal order-up-to policy in an independent and identically distributed (i.i.d.) demand setting, and is a special case of GOUTP. Our key intuitions and conclusions regarding information losses are not restricted by our specific inventory policy structure, they preserve under any affine and stationary ordering policies.

Gaur et al. (2005) and Giloni et al. (2014) study the general ARMA model and conclude that there is no value of information sharing when retailers' demand can be inferred from the order history. We test this condition on our data set and find that our empirical findings contradict with the theoretical predictions in the literature. We recognize a key component

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

absent in the literature: decision makers may deviate from the target policy, whereas the literature assumes that decision makers strictly and consistently follow a replenishment policy. We relax the strict adherence assumption by allowing decision shocks, which successfully explains our substantial empirically evaluated value of information sharing, thus filling the gap between the theoretical and empirical observations.

2.2. Empirical Work

A growing body of empirical literature analyzes the bullwhip effect and information sharing. In a game-theoretic environment, agents have incentives to partially rely on the data or share untruthful information. Cohen et al. (2003) and Terwiesch et al. (2005) empirically show the low efficiency of forecast sharing. Cachon et al. (2007) investigate a wide range of industries and show insignificant variance amplification for some industries. Bray and Mendelson (2012) decompose the bullwhip by information transmission lead time and show a significant amplification generated from last-minute shocks. Using an econometric model, Dong et al. (2014) find that the inventory decision-making transfer between firms, which means the supplier manages the retailer's inventory, benefit both upstream and downstream firms. They show a negative relation between the decision transfer and distributor's average inventory. In our paper, the retailer's demand information is an additional indicator included to help forecast supplier orders. Similarly, one can use other potential indicators to predict customer demand, e.g., financial market index or accounting variables (see Osadchiy et al. 2013, Kesavan et al. 2010, among others).

3. Model

We consider a two-echelon supply chain with a supplier and a retailer. The retailer faces demand D_t and places an order O_t to the supplier in each week t. In each week, the supplier predicts the future order, e.g., the one-step prediction for week t given the history through week $t-1$, which we denote as $\hat{O}_{t-1,\,t}$ (throughout the paper, hats denote forecasted quantities). The supplier aims to improve the forecast accuracy of future orders by including downstream sales data. Without information sharing, the supplier only observes the retailer's order history. With information sharing, in addition to orders, the supplier also observes the retailer's sales history.

Within each week, the following sequence of events occur: (1) the retailer's demand is realized and the retailer places an order to the supplier; (2) after receiving the order, the supplier releases the shipment; (3) the supplier collects the latest information and predicts the future h-step ahead orders; and (4) based on the updated prediction, the supplier makes production decisions.

3.1. Demand Process

We study a similar demand structure as that of Gaur et al. (2005) and Giloni et al. (2014). During each week t, the retailer faces external demand, D_t, for a single item. Let D_t follow an ARIMA(p, d, q) process, where p, d, and q are nonnegative integers that represent the degree of the autoregressive, integrated, and moving average parts of the model, respectively. The ARIMA structure assumes a linear combination of historical observations and historical shocks. When $d = 0$, the ARIMA$(p, 0, q)$ process is reduced to an ARMA(p, q) process,

$$D_t = \mu + \rho_1 D_{t-1} + \rho_2 D_{t-2} + \cdots + \rho_p D_{t-p} + \epsilon_t - \lambda_1 \epsilon_{t-1} - \lambda_2 \epsilon_{t-2} - \cdots - \lambda_q \epsilon_{t-q}, \qquad (1)$$

where μ is the process mean, ϵ_t is an i.i.d. normal demand shock with zero mean and variance σ_ϵ^2, ρ_i is the autoregressive coefficient, and λ_i is the moving average coefficient.

To derive the abbreviated expression for $d \geq 0$, we introduce the backward shift operator B, which shifts variables backward in time; e.g., $B^d D_t$ shifts demand back by d times $B^d D_t = D_{t-d}$, and $(1-B)D_t$ differences demand once $(1-B)D_t = D_t - D_{t-1}$. Differencing the demand twice means differencing $D_t - D_{t-1}$ one more time $(1-B)^2 D_t = D_t - 2D_{t-1} + D_{t-2}$. Similarly, $(1-B)^d D_t$ differences the demand d times, and we refer to it as the dth-order differenced demand.

Let the AR coefficient be denoted as $\phi_{\mathrm{AR}}(B) = 1 - \rho_1 B - \rho_2 B^2 - \cdots - \rho_p B^p$, the integration coefficient as $\pi(B) = (1-B)^d$, and the MA coefficient as $\varphi_{\mathrm{MA}}(B) = 1 - \lambda_1 B - \lambda_2 B^2 - \cdots - \lambda_q B^q$. When $d \geq 0$, $\pi(B)D_t$ follows an ARMA(p, q) process, and we rewrite the demand process (1) as

$$\phi_{\mathrm{AR}}(B)\pi(B)D_t = \mu + \varphi_{\mathrm{MA}}(B)\epsilon_t.$$

We can further rewrite $\pi(B)D_t$ as a moving average (MA) representation with $\varphi(B) = \phi_{\mathrm{AR}}^{-1}(B)\varphi_{\mathrm{MA}}(B)$:

$$\pi(B)D_t = \mu + \varphi(B)\epsilon_t. \qquad (2)$$

We will work with this MA representation because it is mathematically equivalent to an ARMA model but has a more concise expression. We assume that the mean of demand is constant. Under this assumption, $E[(1-B)^d D_t] = 0$ for $d > 0$, and thus, the differenced demand has zero process mean for $d > 0$.

We review a basic, yet important, property of an MA process from the time-series literature: covariance stationarity. For details, we refer readers to Hamilton (1994) and Brockwell and Davis (2002). We assume the dth differenced demand is covariance stationary; that is, the differenced demand has a finite and constant mean, finite variance, and time invariant covariance of $\pi(B)D_t$ and $\pi(B)D_{t+h}$ for any t and h. One might

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

think that the MA model is restricted to a convenient class of models. However, representation (2) is fundamental for any covariance stationary time series. That is, any covariance stationary process is equivalent to an MA process in terms of the same covariance matrix (Wold 1938). Therefore, assuming the ARIMA model is not restrictive. We adopt the Hamilton (1994, p. 109) description of the equivalence between the stationarity and MA representation, which is known as the Wold decomposition property.

Property 1 (Wold Decomposition). *Any zero-mean covariance stationary process X_t can be represented in the MA form $X_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i}$, where $\alpha_o = 1$ and $\sum_{i=0i}^{\infty} \alpha^2 < \infty$. The term ϵ_t is white noise and represents the error in forecasting*: $\epsilon_t \equiv X_t - \hat{E}(X_t \mid X_{t-1}, X_{t-2}, \ldots)$.

3.2. Replenishment Policy

This section presents the replenishment policy the retailer adopts in practice. We interviewed the planner who places orders to understand the policy. We learned that the supplier is the retailer's only source, and that it requires a transportation lead time L_R to ship products to the retailer.

According to the planner, the retailer aims at keeping a constant DOI (days of inventory) amount of the total on-hand inventory and in-transit inventory. In addition, the decision maker smooths orders. (We find this in the data and confirm this with the planner.) We refer to such policy as the "ConDOI policy with order smoothing," where "Con" represents constant and "DOI" represents days of inventory. We first define the ConDOI policy and then extend it with order smoothing.

If a retailer follows the ConDOI policy, she places an order at the end of week t to bring the inventory level up to the target days of inventory multiplied by the retailer's total future demand forecast within the transportation lead time L_R. For example, if the retailer targets the inventory at 14 days and the lead time L_R is 3 weeks, the retailer follows an order-up-to policy with order-up-to level equal to 2 (weeks) × the retailer's demand forecast of the next three weeks. When the demand is i.i.d. distributed, the retailer's demand forecast is constant. Thus, both the optimal order-up-to policy and the ConDOI policy generate constant orders, and they are equivalent under i.i.d. demand. When demands are correlated, the optimal order-up-to level changes every week, which is not convenient from a practical perspective. The ConDOI policy, on the other hand, requires only one parameter, the DOI level, to manage inventory, and thus, is easy to implement and free from heavy computational burdens, which explains its use in practice.

Since a week has seven days, the target weeks of inventory equals 7^{-1} × target DOI level. We denote it as Γ, where Γ is positive and constant.[1] Let $\hat{D}^R_{t,\,t+k}$ denote the retailer's future demand forecast for week $t+k$ made in week t. According to the interview with the retailer, we learned that the retailer adopts a weighted moving average demand forecast, which uses recent demands in H weeks (see Chen et al. 2000b for the moving average forecast and Chen et al. 2000a for the exponential smoothing forecast). Chen et al. (2000b) show that the moving average is one of the most commonly used forecasting techniques in practice. Let $\hat{m}_t$ denote the retailer's forecast of future L_R week demands given demand history prior to week t,

$$\hat{m}_t \equiv \sum_{k=1}^{L_R} \hat{D}^R_{t,\,t+k} = \sum_{j=0}^{H} \beta_j D_{t-j}, \tag{3}$$

where β_j is the coefficient of demand in the past jth weeks, the sum of which equals the transportation lead time, $\sum_{j=0}^{H} \beta_j = L_R$. At the end of week t, the retailer orders up to $\Gamma \hat{m}_t$. Note that the retailer that we study adopts a suboptimal demand forecast[2] (the optimal demand forecast should follow the ARIMA structure). We will argue that this assumption does not have a qualitative impact on the theoretical results in §6.

According to the planner, the ending inventory in each week might not reach the target days of inventory because the order decision might fail to adjust the end inventory changes completely. To address this, we extend the ConDOI policy by allowing a fixed proportion of last week's inventory to become the current week's inventory. In other words, the order-up-to level is a convex combination of that of the ConDOI policy and that of the demand replacement policy,

$$I_t = \gamma \Gamma \hat{m}_t + (1-\gamma) I_{t-1},$$

where γ is the order smoothing level, and it is between $[0, 1]$. Irvine (1981) introduces a similar notion and empirically confirms that firms attempt a partial adjustment toward the optimum level.

Given the fundamental law of material conservation, $O_t = D_t + I_t - I_{t-1}$, we write the order as

$$O_t = D_t + \gamma(\Gamma \hat{m}_t - I_{t-1}). \tag{4}$$

The order in week t is the current week's demand plus γ fraction of the net inventory under the ConDOI policy. The larger γ, the faster the order adjusts to

[1] In the following empirical analysis, we allow Γ to vary across the summer and the winter. Γ remains constant across seasons. Therefore, assuming a constant ConDOI for our theoretical analysis is not restrictive.

[2] To empirically recover the replenishment policy parameters later in the paper, one has to use the same policy that generated the data.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

the target ConDOI inventory level. The order smoothing component enables the extension of the ConDOI policy to a rich family of linear policies. The ordering rule is reduced to a ConDOI policy when $\gamma = 1$, and becomes a demand replacement policy when $\gamma = 0$. The order expression of the myopic order-up-to policy is a special case when $\gamma = 1$. Graves et al. (1998) and Chen and Lee (2009) derive general production-smoothing policies by imposing weight on forecast revisions, which bear an affine and time-invariant structure on historical signals. Our smoothing level γ smooths the net inventory to produce the desired order-up-to level, which is a special case of the above general class. The order O_t in Equation (4) may be negative, in which case we assume that this excess inventory is returned without cost (for the same assumption, see Chen et al. 2000b, Gaur et al. 2005, Chen and Lee 2009).

We can iteratively replace I_{t-i} with $\gamma\Gamma\hat{m}_{t-i} + (1-\gamma)I_{t-i-1}$ for any $i \geq 1$ in the order process (4), which becomes

$$O_t = D_t + \gamma\sum_{i=0}^{H}\Gamma\beta_i D_{t-i} - \gamma^2\sum_{i=1}^{\infty}(1-\gamma)^{i-1}\sum_{j=0}^{H}\Gamma\beta_j D_{t-i-j}.$$

We define

$$\psi(B) \equiv 1 + \gamma\sum_{i=0}^{H}\Gamma\beta_i B^i - \gamma^2\sum_{i=1}^{\infty}\sum_{j=0}^{H}(1-\gamma)^{i-1}\Gamma\beta_j B^{i+j}$$

as the policy parameter. Using $\psi(B)$, we abbreviate the order as $O_t = \psi(B)D_t$. Since demand satisfies $\pi(B)D_t = \mu + \varphi(B)\epsilon_t$ and the order satisfies $\pi(B)O_t = \pi(B)\psi(B)D_t$, we can represent the order process as an ARIMA model with the white noise series $\{\epsilon_t\}$:

$$\pi(B)O_t = \mu + \varphi(B)\psi(B)\epsilon_t, \tag{5}$$

which has the same expression in Gaur et al. (2005, (7)): orders are linear in demand shocks.

The coefficient of ϵ_t in (5) is $c_0 \equiv 1 + \gamma a_0$. Since this coefficient is rarely zero according to our data set, we assume that $c_0 \neq 0$ and normalize it to one (see Giloni et al. 2014 for discussion on $c_0 = 0$). Then the centered order follows an MA process with the white noise series $\{c_0\epsilon_t\}$,

$$\pi(B)O_t - \mu = c_0^{-1}\varphi(B)\psi(B)c_0\epsilon_t. \tag{6}$$

We next empirically test the theoretical model and evaluate the forecast accuracy improvement when there is information sharing.

4. Empirical Estimation

This section sets up the forecasting procedure, explains the data set, and presents the empirical models.

We compare the forecast accuracy under two settings: *NoInfoSharing* and *InfoSharing*. The supplier observes the retailer's order history under the NoInfoSharing setting and observes the additional retailer's sales history under the InfoSharing setting.

We choose the last 26 weeks in our data as the out-of-sample test period. This out-of-sample comparison is made in two stages. First, we forecast the one-step-ahead order over the out-of-sample test period. To be specific, the forecast begins 26 weeks before the end of the data. Given information history through the end of week $t-1$, we predict the order for week t. We then update the information history from the beginning of the data through the end of week t to predict for week $t+1$. We update the available information history on a rolling basis to obtain the order forecast and calculate the forecast error by comparing the actual observation and predicted value. Second, we conduct tests of equal forecast accuracy on the two sequences of forecast errors generated from two candidate forecasting methods.

4.1. Data

We obtain the data from a CPG company, which is a leading manufacturer and supplier in the U.S. beverage and snack food industry. We study two brands of products: sports drinks and orange juice.

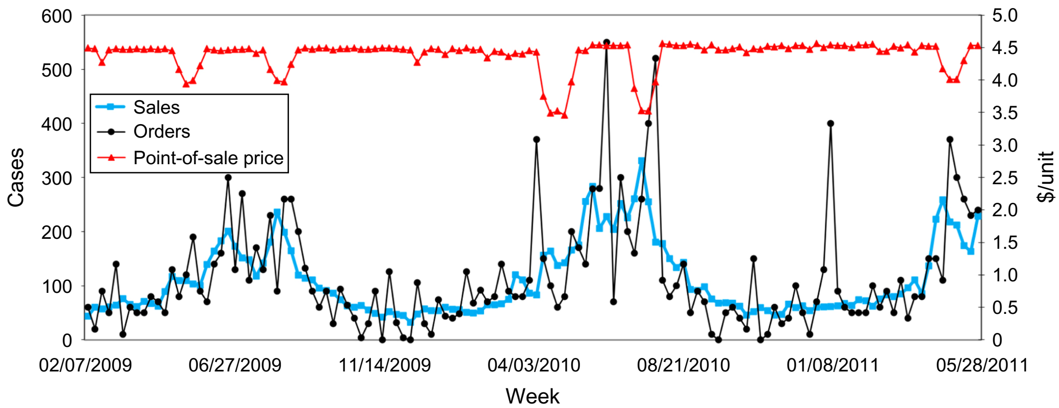
Our data set consists of three elements of a specific retail customer: (1) the retailer's sales from its distribution centers to local stores, (2) the retailer's orders from the retailer's distribution centers to the supplier's distribution center, and (3) the products' retail price. This retail customer is one of the CPG company's major accounts. The data spans 126 weeks between 2009 and 2011. We calculate the retailer's inventory using the fundamental law of material conservation, given sales and orders. The retailer's inventory level stays positive over all weeks, indicating that stock outs are very rare in our data set. In addition, we find in a numerical study that the value of demand and the value of sales are statistically indistinguishable under parameters that are representative of our data set. Thus, we approximate sales as actual demand in our study.

We eliminate untrustworthy data, such as new-entering products that have incomplete data points or obsolete products that are existing the market. After cleaning the data, we have 51 product lines in total: 19 orange juice products and 32 sports drink products. We summarize the sales, orders, and price of a specific product (orange flavor powder product) over 126 weeks in Figure 1. It shows the bullwhip effect: the upstream order has larger volatility than the downstream sales. Further, when there is a price promotion, demand and orders experience a spike during the discount activity and suffer a slump when price returns to normal.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Figure 1 (Color online) Summary of Sales, Orders, and Point-of-Sale Price for Product PD OR



The two major concerns regarding high-promotional products are (1) the violation of the stationary assumption of both demand and orders, and (2) the complicated ordering rules that preclude us from completely removing the promotional lift and correctly estimating the policy parameters.[3] Further, the orders might be moved from peak to nonpeak periods according to certain rules, such as shipping certain percentage of orders in advance prior to the promotional week, if planners anticipate a spike in future demand (Van Donselaar et al. 2010 also show that advancing orders are an important consideration of decision makers). Price variations across weeks generate nonstationary spikes, which further complicates the replenishment policy and results in a time-variant covariance matrix of orders.

For the purpose of our study, we shall classify the products into low-promotional and high-promotional products and focus on the former. This classification is based on the price discount and frequency of discount being offered on the product.[4] Based on the data set, 14 products have low-promotional activities. The 14 low-promotional items occupy 20% of the total ordering volume of the retailer. For completeness, we empirically study the rest of the high-promotional products (see the technical companion, available as supplemental material at http://dx.doi.org/10.1287/mnsc.2014.2132), and we briefly show and discuss the results in §9.

To summarize, we utilize the retailer's (1) sales to the end customers and (2) order fulfillment to the supplier. The summary statistics over the 126 weeks are presented in Table 1. Sales have approximately the same mean as orders because inflows balance outflows. As a side note, we also observe the bullwhip effect: products have higher variations in upstream orders.[5]

4.2. The Empirical Model with Information Sharing: Decision Deviations

We next address the key difference between a theoretical model in the literature and what happens in practice, and explain the *InfoSharing* forecasting method.

The key underlying assumption in the theoretical model described above, as well as in the literature, is that the decision maker strictly and consistently follows a specific replenishment rule. This is rarely the case in practice, because decision makers only use the replenishment policy as a mere guide from which they rationally deviate. Van Donselaar et al. (2010) show that retail store managers may not follow order advices by the replenishment system because their incentives may differ from the system or they perceive the system to be suboptimal. We interviewed planners that place orders to the CPG company, and we observed how they place orders. The observations confirm the literature that the planners' orders indeed deviate from the suggested policy. Deviations could stem from several operational causes. For example, to increase transportation efficiency and reduce transportation cost, the retailer tries to fill up a full truck, and thus she may place a rounded order quantity or

[3] To be specific, when retailers forecast future demand of promotional products in practice, they first generate the baseline (nonpromotional) forecast as in Equation (3) and then add back the promotional lift by multiplying the price reduction rate.

[4] Negotiated at the beginning of each year, the supplier has a fixed price plan throughout the year. Thus, the future price can help predict demand changes for promotional products. A promotional activity can last for several weeks. We define a promotional depth metric to capture price discount and frequency. Promotional depth sums every promotional activity's price discount measured as a percentage within the test periods, $\sum_i$ discount rate$_i$, where i is the total number of activities in the test periods. We define the low-promotional products as those with promotional depth ≤ 0.15 (zero or one promotional activity), and we define the high-promotional product as those with higher promotional depth.

[5] The literature discusses the potential factors of the bullwhip effect, such as the inventory policy and forecasting technique (Lee et al. 1997; Chen et al. 2000a, b), order batching (Lee et al. 1997), and order smoothing (Chen and Lee 2009).

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Table 1 Summary Statistics of Sales and Orders

		Sales			Orders		
Brand	Product	Mean	S.D.	C.V.	Mean	S.D.	C.V.
Orange juice	128 OR	1,387.41	395.32	0.28	1,383.64	802.82	0.58
	128 ORCA	1,019.71	290.39	0.28	1,028.06	582.27	0.57
	12 OR	457.83	129.67	0.28	456.59	263.77	0.58
	12 ORCA	227.47	84.79	0.37	233.24	164.90	0.71
	59 ORST	593.18	161.12	0.27	601.90	281.29	0.47
	59 ORPC	724.95	188.55	0.26	727.02	346.00	0.48
Sports drink	500 BR	870.48	391.70	0.45	889.03	726.24	0.82
	500 GP	1,145.93	474.19	0.41	1,171.95	851.16	0.73
	PD LL	75.33	42.28	0.56	78.10	74.87	0.96
	PD OR	86.74	51.74	0.60	90.74	82.20	0.91
	PD FRZ	97.84	53.96	0.55	100.91	83.25	0.82
	1GAL GLC	511.20	95.10	0.19	517.38	199.86	0.39
	1GAL FRT	336.57	83.80	0.25	340.76	141.59	0.42
	1GAL OR	522.55	95.43	0.18	531.03	201.29	0.38

Note. C.V., coefficient of variation.

more (or less) orders than the policy suggests. Products with inventory above the target DOI level might still be replenished because as the week approaches Friday, the decision makers overreplenish to guarantee enough inventory during the weekends. In practice, the retailer might place orders daily. However, for this study, we have access only to the weekly level instead of daily information. Looking through the lens of the aggregate data, we lose the detail on the replenishment decision, which is reflected by the actual orders' deviations from the theory.

Among the above different operational drivers, a common characteristic is that they can be observed by decision makers, but not by statisticians. Henceforth, we rationalize the retailer's departure from the exact policy following the same spirit as Rust (1994): it is due to a state variable that is observed by the downstream retailers but not by the upstream suppliers. It is interesting to note that deviations from the prescribed order quantity might further reduce system cost since they are based on more detailed information.

We extend the theoretical framework by explicitly including such idiosyncratic shocks in decision making. We refer to such idiosyncratic shocks as decision deviations. We assume the decision deviation δ_t is normally distributed with zero mean and variance σ_δ^2, and independent with historical demand shock $\epsilon_s, s < t$. However, contemporaneous demand signals and decision deviation signals can be correlated. A common approach in the empirical literature is to model this error term as additively separable in the decision. Using this approach, we obtain

$$O_t = D_t + \gamma(\Gamma\hat{m}_t - I_{t-1}) + \delta_t. \tag{7}$$

We shall show that the inclusion of this zero-mean shock in the theoretical model has important consequences on the value of information sharing in §6.

As before, we iteratively replace I_{t-i} with $\gamma\Gamma\hat{m}_{t-i} + (1-\gamma)I_{t-i-1} + \delta_{t-i}$ in the order process (7) and obtain

$$O_t = D_t + \gamma\sum_{i=0}^{H} a_i D_{t-i} - \gamma^2 \sum_{i=1}^{\infty}(1-\gamma)^{i-1}\sum_{j=0}^{H} a_j D_{t-i-j} + \delta_t - \sum_{i=1}^{\infty}\gamma(1-\gamma)^{i-1}\delta_{t-i}.$$

We define $\kappa(B) = 1 - \gamma\sum_{i=1}^{\infty}(1-\gamma)^{i-1}B^i$ as the order smoothing parameter. Applying the backshift operators $\kappa(B)$ and $\pi(B)$, the order process can be abbreviated as

$$\pi(B)O_t - \mu = \varphi(B)\psi(B)\epsilon_t + \pi(B)\kappa(B)\delta_t. \tag{8}$$

Remark. Note that decision deviations focus on how much real order decisions depart from the replenishment policy that the firm is expected to follow. Although decision deviations and the bullwhip effect share similar operational factors, these two concepts are different. For instance, decision deviations can be present in the absence of the bullwhip effect: if a firm orders a fixed amount (truck load effect) each week but is expected to follow the ConDOI policy, the order variance is zero (no bullwhip effect), but the firm deviates from the target inventory policy (positive decision deviations). On the contrary, the bullwhip effect can be present in the absence of decision deviations. For example, when a firm strictly follows a general-order-up-to policy, decision deviations are zero, whereas the order smoothing level still amplifies the order variability (Chen and Lee 2009). They may also coincide: under the demand replacement policy with $O_t = D_t + \delta_t$, decision shocks are reflected in δ_t, which also directly drives a higher upstream variance.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

InfoSharing Method. In practice, the retailer's forecast is a weighted summation of last month's demands. Thus, we let $H = 3$ and the order in (7) becomes $O_t = (1+\gamma\Gamma\beta_0)D_t + \gamma\Gamma\beta_1 D_{t-1} + \gamma\Gamma\beta_2 D_{t-2} + \gamma\Gamma\beta_3 D_{t-3} - \gamma I_{t-1} + \delta$. We estimate the replenishment policy parameters for each week in the test period by

$$O_t = c_0 D_t + c_1 D_{t-1} + c_2 D_{t-2} + c_3 D_{t-3} - \gamma I_{t-1} + \delta_t, \qquad (9)$$

where $c_0 \equiv 1+\gamma\Gamma\beta_0$ and $c_i \equiv \gamma\Gamma\beta_i$ for $i = 1, 2, 3$. The idiosyncratic shock, δ_t, captures the decision maker's deviation from the deterministic replenishment policy.[6] If δ_t is positive, the retailer orders more than what our policy predicts, and vice versa. As we will show later, δ_t is the key element in bridging the empirical and theoretical results.

To forecast the supplier's order in $t+1$, we first forecast future demands. We fit the ARIMA model on the historical demand series to forecast $\hat{D}_{t,t+1}$. We obtain the best estimator with the step-wise variable selection method, which chooses the model with the lowest Bayesian information criterion (BIC).[7] The order prediction for $t+1$ uses the parameters estimated from (9), $\hat{D}_{t,t+1}$, D_s, where $s \le t$ and I_t, $\hat{O}_{t,t+1} = c_0\hat{D}_{t,t+1} + c_1 D_t + c_2 D_{t-1} + c_3 D_{t-2} - \gamma I_t$. Note that $\hat{D}_{t,t+1}$ is an optimal demand forecast, which differs from the retailer's weighted average demand forecast $\hat{D}^R_{t,t+1}$.

4.3. The Empirical Model Without Information Sharing

For the NoInfoSharing benchmark, we use only the order history to predict future orders. We replicate the CPG company's current practice: using the ARIMA process to model orders and make future predictions. We next show that this method is also theoretically grounded.

ARMA-In-ARMA-Out Property. The order process with decision deviations has a stationary covariance. According to property 1, the order process (8) follows an ARIMA model. This is consistent with the "ARMA-in-ARMA-out" (AIAO) property discussed in the literature (Gilbert 2005, Gaur et al. 2005), where AIAO means that the retailer's order process is also an ARMA process with respect to the demand shock.[8] If the replenishment policy is an affine and time invariant function of the historical demand, inventory, demand shock, and decision deviations, the order process has a stationary covariance. Henceforth, the upstream order series also follows an ARIMA process under such policies.

NoInfoSharing Method. Based on the above AIAO property, we fit an ARIMA(p,d,q) model to the order history,

$$(1-B)^d O_t = \mu + \tilde{\rho}_1(1-B)^d O_{t-1} + \cdots + \tilde{\rho}_p(1-B)^d O_{t-p} + \eta_t + \tilde{\lambda}_1\eta_{t-1} + \cdots + \tilde{\lambda}_q\eta_{t-q}, \qquad (10)$$

where η_t is the order shock, and μ, $\tilde{\rho}_i$ and $\tilde{\lambda}_i$ are the time-series coefficients of the order process. As before, we predict future orders by applying the estimated ARIMA model. For example, if $d = 1$ and we have the available information history until week $t-1$, then the order forecast for week t uses estimated coefficients μ, $\tilde{\rho}_i$, and $\tilde{\lambda}_i$, and estimated historical order shocks $\eta_{t-1}, \eta_{t-2}, \ldots$, $\hat{O}_{t-1,t} = O_{t-1} + \mu + \tilde{\rho}_1(O_{t-1} - O_{t-2}) + \tilde{\rho}_2(O_{t-2} - O_{t-3}) + \cdots + \tilde{\rho}_p(O_{t-p} - O_{t-p-1}) + \tilde{\lambda}_1\eta_{t-1} + \tilde{\lambda}_2\eta_{t-2} + \cdots + \tilde{\lambda}_q\eta_{t-q}$. This method is a reliable representation of the CPG company's current practice.[9]

5. Empirical Results

This section provides empirical evidence that incorporating downstream sales data improves order forecast accuracy, compared to the benchmark where sales information is not shared. We display the estimated demand and replenishment policy parameters and the empirical findings.

5.1. Parameter Results

We present the demand and replenishment policy parameters in Table 2. The first column records the (p,d,q) value of the ARIMA demand, and the next three columns are the corresponding demand parameters; i.e., the 128 OR product follows an ARIMA$(0,1,1)$ demand process, $D_t = D_{t-1} + \epsilon_t - 0.93\epsilon_{t-1}$. For all products, demand is best estimated by $d = 1$, which implies that the first-order differenced demand is an ARMA process. The estimated demand parameters are used to generate the optimal demand forecast $\hat{D}_{t,t+1}$ for the InfoSharing forecast. In the policy

[6] The estimating Equation (9) does not have an intercept, which might result in a nonzero average of δ_t in the estimation. Note that we empirically test that for most products, δ_t has a zero mean ($p < 0.05$).

[7] BIC is a criterion for model section for time-series analysis and model regression. It selects the set of parameters that maximizes the likelihood function with the least number of parameters in the model.

[8] Giloni et al. (2014) show a more general QUARMA-in-QUARMA-out property that includes the special case $c_0 = 0$.

[9] It is important to ensure that the information environment and methods used for our paper and the company are similar, so that the comparison of forecast accuracy is fair. We conduct comparisons between the company's historical forecasts and our derived forecasts for promotional and nonpromotional products, at the stock keeping unit (SKU) and pack level, and across winter periods and summer periods. The forecast accuracy turns out to be very close (around an average 2% difference) and the company confirmed it as insignificant.

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Table 2 Estimated Demand and Policy Parameters

		Demand parameters				Policy parameters						
Brand	Product	(p, d, q)	λ_1	λ_2	λ_3	c_0	c_1	c_2	c_3	γ	DOI	Weight
Orange juice	128 OR	(0, 1, 1)	0.93 (0.04)			1.30 (0.18)	0.27 (0.16)			0.63 (0.12)	6.36	0.62
	128 ORCA	(0, 1, 1)	0.93 (0.04)			1.33 (0.18)	0.30 (0.17)			0.53 (0.12)	8.48	0.57
	12 OR	(0, 1, 2)	0.48 (0.10)	0.33 (0.10)		1.46 (0.17)		0.60 (0.18)		0.87 (0.12)	8.56	0.86
	12 ORCA	(0, 1, 2)	0.28 (0.10)	0.25 (0.10)		0.97 (0.21)	1.09 (0.23)		0.36 (0.19)	0.86 (0.12)	11.54	0.85
	59 ORST	(0, 1, 1)	0.72 (0.07)			1.55 (0.10)				0.39 (0.07)	9.76	0.68
	59 ORPC	(0, 1, 1)	0.8 (0.07)			1.84 (0.18)			−0.46 (0.22)	0.28 (0.08)	9.58	0.73
Sports drink	500 BR	(0, 1, 3)	0.12 (0.09)	0.16 (0.09)	0.49 (0.09)	1.17 (0.29)	0.55 (0.33)			0.35 (0.07)	14.50	0.93
	500 GP	(0, 1, 0)				0.39 (0.25)	0.67 (0.38)	0.63 (0.29)		0.36 (0.07)	13.43	0.88
	PD LL	(0, 1, 0)				0.61 (0.40)	1.11 (0.44)			0.24 (0.06)	21.06	0.85
	PD OR	(0, 1, 1)	0.30 (0.10)			1.37 (0.16)		0.39 (0.23)		0.29 (0.07)	18.35	0.84
	PD FRZ	(0, 1, 1)	0.32 (0.09)			1.27 (0.16)		0.47 (0.22)		0.32 (0.07)	16.22	0.83
	1GAL GLC	(0, 1, 2)	0.2 (0.09)	0.43 (0.09)		0.76 (0.25)	1.17 (0.28)		−0.54 (0.21)	0.22 (0.08)	12.38	0.87
	1GAL FRT	(0, 1, 2)	0.37 (0.10)	0.36 (0.10)		1.49 (0.13)				0.25 (0.07)	13.38	0.78
	1GAL OR	(0, 1, 2)	0.29 (0.10)	0.34 (0.10)		1.52 (0.12)				0.30 (0.07)	12.33	0.77

Notes. The number in parentheses denotes the standard error of the estimate. For the policy parameters, we apply the step-wise variable selection method to only include variables with $p < 0.05$ in the regression.

parameter sector, orders are determined by a linear combination of historical demands[10] and the last week's inventory. The order smoothing level γ is statistically significantly different from one, providing strong support for retailers' order smoothing.[11]

In practice, it takes one week to ship products from the CPG company to the retailer. Thus, the transportation lead time is one, $L_R = 1$, and we have

$$\sum_{i=0}^{H} \Gamma \beta_i = 1 \quad \text{and} \quad \Gamma = \gamma^{-1}\left(\sum_{i=0}^{H} c_i - 1\right).$$

The DOI column displays the estimated days of inventory level. For example, the retailer aims to keep 6.36 days (0.91 weeks) of inventory for product 128 OR, which is equivalent to $0.91 \times 1{,}387.41 = 1{,}262.54$ cases of inventory on average. According to the planner, the retailer targets at a lower DOI level for the orange juice brand and a higher DOI for the sports drink brand, because the sports drink products have higher demand variations. Our estimated DOI level is consistent with the actual target level claimed by the decision maker. Decision deviations are the residuals when estimating (9). They satisfy the white noise assumption according to the Bartlett test. We define $\sigma_\delta^2/(\sigma_\epsilon^2 + \sigma_\delta^2)$ as the decision deviation weight, which measures the relative weight of the decision uncertainty over the demand uncertainty. We display the weight in the last column. The weight of δ over

[10] For some products, the estimated weight of the current week's demand is zero, which is unlikely to occur in practice. Our estimation shows a zero coefficient because the retailer may replenish inventory during the week, but our data set consists of system's snapshots at the end of the each week. If the retailer replenishes certain products always on Monday, the current week's order should be a linear combination of past weeks' demand, not including the current week (since the current week's demand has not been realized yet). Therefore, we adjusted for products with zero c_0 in Table 2 by shifting c_1, c_2 and c_3 forward.

[11] The order smoothing level γ is necessary in improving forecasts. We test a setting without allowing order smoothing (setting $\gamma = 1$). When $\gamma = 1$ (this corresponds to the retailer following the strict ConDOI policy) in the estimating equation, the in-sample fit (R^2, AIC and BIC) is low, and we do not observe a significant out-of-sample forecast accuracy improvement.

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Table 3 NoInfoSharing and InfoSharing Forecast Accuracy Comparison in MAPE and RMSE

			MAPE (%)			Relative RMSE (%)		
Brand	Index	Product	NoInfo Sharing	Info Sharing	MAPE % improve	NoInfo Sharing	Info Sharing	MSE % improve
Orange juice	1	128 OR	39.0	21.5	45.0**	47.9	43.3	18.1**
	2	128 ORCA	42.7	29.8	30.3*	45.4	38.9	26.5**
	3	12 OR	198.7	82.2	58.6*	77.8	53.2	53.4**
	4	12 ORCA	106.7	53.1	50.2**	59.3	40.6	53.1**
	5	59 ORST	40.0	32.5	18.8*	48.3	46.6	7.1
	6	59 ORPC	26.3	19.0	27.7**	33.6	28.2	29.4**
Sports drink	7	500 BR	40.0	24.1	39.8**	47.5	29.1	62.5**
	8	500 GP	32.6	20.9	36.0**	34.9	19.6	68.4**
	9	PD LL	25.2	23.9	4.7	43.8	30.6	51.3
	10	PD OR	68.3	38.1	44.2**	81.6	35.5	81.1*
	11	PD FRZ	37.9	22.9	39.5*	41.3	27.1	56.9**
	12	1GAL GLC	37.6	23.3	38.0**	43.1	29.2	54.2**
	13	1GAL FRT	50.9	35.7	29.9**	52.5	35.6	54.0**
	14	1GAL OR	34.4	23.9	30.4*	36.6	27.2	44.8**

Note. InfoSharing forecasts outperform NoInfoSharing forecasts for almost all products.
**At level $p < 0.05$, the accuracy improvement over the NoInfoSharing method is significant.
*At level $p < 0.1$, the accuracy improvement over the NoInfoSharing method is significant.

demand signals ranges from 0.6 to 0.9 across products, providing strong evidence that decision deviations are prevalent in our data set.[12]

5.2. Including Downstream Demand Improves Order Forecasting

We measure the accuracy with three forecast error metrics widely used in the literature (cf. Osadchiy et al. 2013, Kesavan et al. 2010): mean absolute percentage error (MAPE), mean squared error (MSE), and relative root mean error over the mean of orders (relative RMSE). Let N be the number of weeks in the test period. The MAPE metric over the test period is $\text{MAPE} = (1/N)\sum_{i=1}^{N} |O_{t+i} - \hat{O}_{t+i-1,\,t+i}|/O_{t+i}$. MAPE measures the absolute error relative to the mean, which is closely related to the metric used by the company from which we received the data. MSE is a frequently adopted accuracy metric in the theoretical literature because of its mathematical tractability. We also use this metric for our theoretical analysis. We display the RMSE value because it is more intuitive to understand.

Table 3 presents the forecast accuracy for each product in MAPE and RMSE separately. The first sector displays the MAPE metrics of each method and the MAPE percentage improvement of the InfoSharing method over the NoInfoSharing method, which is defined by $(\text{MAPE}_{\text{NoInfo}} - \text{MAPE}_{\text{Info}})/\text{MAPE}_{\text{NoInfo}}$. The larger the percentage improvement, the more accurate the forecast with information sharing. We conduct the pairwise t-test to determine the statistical significance of forecast improvement. A major drawback of MAPE is that zero or small real observations might distort the measure. For example, 12 OR and 12 OR CA have low ordering quantity in three weeks, which leads to a high MAPE even on the average level. The relative RMSE sector reports the RMSE value over the mean of orders. We test the significant level by MSE and display the MSE percentage improvement in the last column, $(\text{MSE}_{\text{NoInfo}} - \text{MSE}_{\text{Info}})/\text{MSE}_{\text{NoInfo}}$.

Table 3 shows that adding downstream demand leads to statistically significant forecast accuracy improvement. For all products (except for PD LL[13]), the InfoSharing method generates statistically significant improvements over the NoInfoSharing method for at least one error metrics. If measuring at the overall level across products, the NoInfoSharing forecasts have a 56.45% MAPE, the number of which is representative of the typical number we observe at the CPG company. At the overall level, the InfoSharing forecasts have a statistically significantly lower MAPE of 33.36%. To summarize, we have empirically showed that the value of downstream demand is statistically significantly positive, and the effect is very large.

[12] The decision deviation δ is important in improving forecasts. We conduct an empirical forecast study excluding decision deviations, which only uses demand parameters and demand signals in Equation (5) to make predictions. The InfoSharing forecasts become very inaccurate for all products, which even perform worse than the NoInfoSharing forecasts.

[13] We also test forecast accuracy using the mean absolute error (MAE) metrics. MAE is defined as $\sum_{i=1}^{N} |O_{t+i} - \hat{O}_{t+i-1,\,t+i}|/\sum_{i=1}^{N} O_{t+i}$. For the product PD LL, the MAE metric shows that the InfoSharing method is statistically significantly ($p < 0.1$) better than the NoInfoSharing method.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

So far, we have empirically tested the value of downstream demand. We next revisit the theoretical conditions in the literature on our products, and compare the theoretical predictions with our empirical observations.

6. Theoretical Results

In this section, we first revisit the main results from Gaur et al. (2005) and Giloni et al. (2014) on the value of sharing information in our settings. We then show the inconsistency between our empirical findings and the corresponding theoretical results suggested in the literature. Finally, we identify the impact of decision deviations on the value of information sharing, and we prove that by incorporating decision deviations, the value of information sharing is always positive if there is uncertainty in both decision deviations and demand.

6.1. Revisit the Literature

When revisiting the literature, we focus on the one-step-ahead forecast and use it as a theoretical foundation to compare with the empirical results. In our new model, we will study the general h-step-ahead forecast.

Besides the covariance stationarity property discussed in §3, we review another important property of a time-series process called invertibility. An MA process is determined by a unique covariance matrix. A covariance stationary process may have multiple MA representations in terms of different sets of coefficients α_i relative to their corresponding white noise series. Among the alternative representations, there is always an invertible representation with respect to some set of shocks, and we are only interested in this invertible one. An MA process $X_t = \mu + \varphi(B)\epsilon_t$ is invertible relative to $\{\epsilon_t\}$ if the shock can be written as an absolutely summable sequence of past demands. A sequence $\{\alpha_t\}$ is said to be absolutely summable if $\lim_{n\to\infty}\sum_{i=0}^{n}|\alpha_i|$ is finite.

Property 2 (Invertibility). *Define $\varphi(z) = 1 - \lambda_1 z^1 - \lambda_2 z^2 - \cdots - \lambda_q z^q$. Then ϵ_t can be written as an absolutely summable series of $\{X_s\}$ with $s \le t$, if and only if all roots of $\varphi(z) = 0$ lie outside of the unit circle, $\{z \in \mathbb{C}, |z| > 1\}$. We say that X_t is invertible relative to $\{\epsilon_t\}$.*

Invertibility guarantees future independence: X_t is only correlated with past value of ϵ_t. Noninvertibility would allow for correlation with future values, which is undesirable. Invertibility is a property of the MA coefficients relative to the corresponding white noise series. According to Brockwell and Davis (2002, p. 54), for any noninvertible process $X_t = \varphi(B)\epsilon_t$, we can find a new white noise sequence $\{w_t\}$ and a new coefficient $\varphi'(B)$ such that $X_t = \varphi'(B)w_t$ and X_t is invertible relative to $\{w_t\}$. We say that the coefficient $\varphi'(B)$ is in the invertible representation. Since empirical estimation identifies parameters based on history, estimators should have invertible representations. Henceforth, we assume the differenced demand process, $(1-B)^d D_t$, satisfies invertibility.

As Hamilton (1994, p. 68) shows, an MA process has at most one invertible representation, which has a larger white noise variance than any other noninvertible representations. Later, we will illustrate that the enlarged white noise caused by converting from the noninvertible to invertible representation is the trigger to the positive value of information sharing in the literature.

In our theoretical analysis, we measure the forecast accuracy by the mean squared forecast error. We denote the information set that contains the historical orders until week t as Ω_t^O and the information set that contains the historical demand until week t as Ω_t^D. The one-step mean squared forecast error without information sharing is $\operatorname{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O)$, and with information sharing is $\operatorname{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O \cup \Omega_t^D)$. The value of information sharing is positive if and only if including downstream information reduces the forecast error

$$\operatorname{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O \cup \Omega_t^D) < \operatorname{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O). \quad (11)$$

Recall the centered order is an MA process $c_0^{-1}\varphi(B)\psi(B)c_0\epsilon_t$. With downstream demand information, the demand and policy parameters can be estimated and thus are known to the supplier. The only uncertainty stems from the demand shock occurring in $t+1$. Thus, we have $\operatorname{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O \cup \Omega_t^D) = \operatorname{Var}(c_0\epsilon_t)$.

Without information sharing, the supplier analyzes the order history as an MA process. If the order has a noninvertible MA representation with respect to demand shocks, i.e., $c_0^{-1}\varphi(B)\psi(B)$ is not invertible, the supplier will not be able to recover demand shocks using the order history. Yet, the supplier will always be able to express the order as an invertible MA representation with respect to some set of shocks. Recall that the shocks in the invertible representation has the largest variance. Therefore, when $c_0^{-1}\varphi(B)\psi(B)$ is not invertible, the variance of the shocks that are invertible is larger than $\operatorname{Var}(c_0\epsilon_t)$. Then inequality (11) holds, and hence the value of information sharing is positive. The positive value of information sharing is equivalent to the noninvertibility property. Gaur et al. (2005) and Giloni et al. (2014) show the same intuitions for the positive value of information sharing. The following proposition gives a formal statement of the sufficient and necessary conditions that sharing downstream demand benefits the supplier's order forecast.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Proposition 1. *If the decision maker strictly adheres to a replenishment policy, the value of information sharing under the one-step forecast lead time is positive if and only if at least one root of $\psi(z)=0$ lies inside the unit circle.*

The value of information sharing is positive if and only if $\varphi(B)\psi(B)$ is in the noninvertible representation, which in turn is equivalent to the existence of at least one root of $\varphi(z)\psi(z)=0$ that lies inside the unit circle. Since all roots of $\varphi(z)=0$ lie outside the unit circle because of the invertible assumption, the order is noninvertible relative to ϵ_t if and only if there exists at least one root of the policy parameter polynomial $\psi(z)=0$ that lies inside the unit circle.

Comparison with Results in the Literature. Recall that we introduce decision deviations when conducting empirical forecasts, which capture idiosyncratic shocks in ordering decisions due to private information observed by retailers. Decision deviations allow us to relax the strict adherence to replenishment policies assumption, and they are the major difference between the model in the literature and our empirical model. We use the following analysis to illustrate that when decision deviations are absent, the theoretical results are different from empirical observations.

We check the invertibility condition in Proposition 1 on the estimated replenishment policy parameters in Table 2. The replenishment policy parameters[14] are invertible for over half of the products. Consider, for example, product 128 ORCA with $c_0=1.33$, $c_1=0.3$, and $\gamma=0.53$. Its policy parameter is

$$\begin{aligned}\psi(z) &= c_0+(c_1-\gamma(c_0-1))z\\ &\quad -\gamma(c_1+(1-\gamma)(c_0-1))\sum_{i=2}^{\infty}(1-\gamma)^{i-2}z^i\\ &= 1.33+0.13z-0.24\sum_{i=2}^{\infty}0.47^{i-2}z^i.\end{aligned}$$

We have $\psi(1)=1.33+0.13-0.24/(1-0.47)>0$. Since $\psi(z)=1.33+z(0.13-0.24/(1/z-0.47))$, we have $\psi(z)>0$ for $0<z<1$. No root of $\psi(z)=0$ lies inside of the unit circle. This means that orders are a sufficient statistic of demand for this product, which allows us to recover the exact demand series from the order history. Thus, the theory shows zero value of information sharing for product 128 ORCA. We find that 10 out of 14 products have invertible policy coefficients, suggesting that these 10 products can gain nothing from information sharing. This is contrary to the empirical observations that show significant improvements from incorporating downstream demand for almost all products. Such different results call for a better theoretical understanding of how decision deviations can alter the result on the value of information.

[14] Note that the replenishment policy parameters are estimated using both orders and demand. Measuring the invertibility of the policy parameters is equivalent to measuring whether orders are invertible to demand signals.

6.2. Preliminary Results

Recall that when decision deviations are present, the order process, after being integrated, consists of two MA processes with demand signals and decision deviations as their corresponding white noise series in (8). Before we delve into the analysis of our specific model, we first study a more general setting: forecasting the aggregation of multiple MA processes. We then apply the results to our model.

Consider N processes $X_t^i=\chi_i(B)\epsilon_t^i$, each of which follow a MA structure with respect to i.i.d. random shock ϵ_t^i. The coefficient is $\chi_i(B)=1+\lambda_1^iB+\lambda_2^iB^2+\cdots+\lambda_{q_i}^iB^{q_i}$ with degree q_i (where q_i can be infinite). Note that we do not impose any restrictions on the coefficient—$\chi_i(B)$ can be either invertible or noninvertible. When predicting future value beyond q_i periods, the forecast is constant and uncertainty cannot be resolved. We allow contemporaneous signals to be correlated, but require signals to be independent across periods. That is, ϵ_t^i is independent of ϵ_s^j for any $s<t$. We also require that the contemporaneous signals are linear independent in Assumption A1. To be specific, the signal of any process is not linear in that of other processes.

Assumption A1. *ϵ_t^i is not a linear combination of ϵ_t^{-i} for any i, where $^{-i}$ represents other processes except i.*

The summation of N processes is $S_t=\sum_{i=1}^N X_t^i=\sum_{i=1}^N \chi_i(B)\epsilon_t^i$. According to Brockwell and Davis (2002, p. 54), S_t has a unique invertible MA representation, which we denote as $\chi_S(B)\eta_t$ with respect to shocks η_t. The coefficient is $\chi_S(B)=1+\theta_1B+\theta_2B^2+\cdots+\theta_{q_S}B^{q_S}$, and q_S is the largest k that guarantees nonzero covariance $\mathrm{Cov}(S_t,S_{t+k})\neq 0$. We impose a technical assumption that $\chi_S(B)$ is not a common factor of all $\chi_i(B)$ in Assumption A2. Then, there exists a process k such that $\chi_S(B)$ is not a factor of $\chi_k(B)$, which guarantees that $\chi_S^{-1}(B)\chi_k(B)$ is of infinite degree.

Assumption A2. *There exists a process k such that $\chi_S^{-1}(B)\chi_k(B)$ is of infinite degree.*

With full information, we not only observe each process but also know each process's coefficient $\chi_i(B)$ and white noise series ϵ_t^i. Specifically, in our setting, with full information, the supplier knows both demand history and knowledge of the retailer's replenishment policy. The coefficients $\chi_i(B)$ consist of demand parameters and replenishment policy parameters, which can be estimated out using historical demand and the replenishment policy. With aggregate information, we

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

only observe the aggregate process S_t. The value is positive for the h-step-ahead forecast if and only if

$$\mathrm{Var}\left(\sum_{l=1}^{h}(S_{t+l}-\hat{S}_{t,\,t+l})\,\Big|\,\bigcup_i \Omega_t^{X^i}\right) < \mathrm{Var}\left(\sum_{l=1}^{h}(S_{t+l}-\hat{S}_{t,\,t+l})\,\Big|\,\Omega_t^{S}\right),$$

where $\hat{S}_{t,\,t+l}\,|\,\bigcup_i \Omega_t^{X^i}$ denotes the best linear forecast under information sharing where the supplier uses both the knowledge of the replenishment policy and the historical demand to obtain demand shocks and decision deviations, and $\hat{S}_{t,\,t+l}\,|\,\Omega_t^S$ denotes the best linear forecast under no information sharing where the supplier uses the historical orders to obtain order shocks. The following theorem states the sufficient condition for the positive value of information sharing.

Theorem 2. *Under Assumptions* A1 *and* A2, *if there exist two processes with different coefficients*, $\chi_i(B)\neq\chi_j(B)$ *for some* i, j, *then*

$$\mathrm{Var}\left(\sum_{l=1}^{h}(S_{t+l}-\hat{S}_{t,\,t+l})\,\Big|\,\bigcup_i \Omega_t^{X^i}\right) < \mathrm{Var}\left(\sum_{l=1}^{h}(S_{t+l}-\hat{S}_{t,\,t+l})\,\Big|\,\Omega_t^{S}\right)$$

for any finite forecast lead time h, *where* $h\le\max_i\{q_i\}$.

Among N processes, if the coefficients of any two processes differ, the aggregate process has a strictly larger mean squared forecast error as long as the forecast is within the effective forecast range.[15] Intuitively, the error variance is different for h if the h-step-ahead forecasts with and without information sharing are different. This theorem indicates that if signals do not evolve in the same manner over aggregation, we can forecast better by distinguishing the signal series and analyzing them separately.

Remark. When the coefficients are the same, the aggregate process becomes $\chi_1(B)\sum_{i=1}^{N}\epsilon_t^i$. We can apply the result of Proposition 1 that the value of information sharing is determined by the invertibility of $\chi_1(B)$. When Assumption A1 is violated, there could be no value of information sharing even when the coefficients are different.[16] When Assumption A2 is violated, there could be no value of information sharing even when the coefficients are different.[17] Note that we relax Assumption A2 and show the extension of Theorem 2 in the technical companion.

6.3. Strictly Positive Value of Information Sharing

Let us apply this general result to the order process in (8) in our setting. The two MA series that constitute the order process are demand shock series and decision deviation series,

$$X_t^1 = c_0^{-1}\varphi(B)\psi(B)c_0\epsilon_t \quad\text{and}\quad X_t^2=\pi(B)\kappa(B)\delta_t. \qquad (12)$$

Based on the Pearson correlation test on demand and decision deviations in our data set, we find low and insignificant correlations when the signals are from the same period or across different periods, which suggests that Assumption A1 holds for our data set and the independence assumption across noncontemporaneous signals holds. Assumption A2 also holds for (12).[18] Sharing demand information corresponds to full information defined in the preliminary analysis, since we can estimate out all the coefficients, demand signals, and decision deviations. No information sharing corresponds to aggregate information, since we only observe the order history. The following proposition illustrates the result of the positive value of information sharing.

Proposition 3. *Given that demand signals and decision deviations are nonzero*, *the value of information sharing is strictly positive if* (a) *demand follows* ARMA(p,q) *for any* h*-step-ahead forecast*, *where* $h\le\max\{q_\epsilon,q_\delta\}$, *or* (b) *demand follows* ARIMA(p,d,q) *for the one-step-ahead forecast*.

Recall that Theorem 2 is based on stationary processes. Under condition (a), demand is stationary (i.e., $d=0$), and we can directly apply Theorem 2 to show that the value of information sharing is positive for *any* h-step-ahead forecast. Under condition (b), demand is nonstationary (i.e., $d\ge1$), which stems from the integrated inclusion of the past observations. Note that the forecast error of the one-step-ahead forecast includes only shocks (not the historical observations). Since the shocks of an ARIMA process are stationary, the arguments used for condition (a) will

[15] If $h\ge\max_i\{q_i\}$, the order forecast becomes the mean of orders with and without downstream information, and thus, there is no value of information sharing. If $q_i=0$ for all i, all processes become an i.i.d. normal process. The order forecast is the mean of orders, and thus the value of information sharing is zero.

[16] For example, consider two processes with the same signals, $X_t^1=\epsilon_t-0.5\epsilon_{t-1}$ and $X_t^2=\epsilon_t-0.7\epsilon_{t-1}$. The summation is $S_t=(2\epsilon_t)-0.6(2\epsilon_{t-1})$ with signal $2\epsilon_t$. Since $\chi_S(B)=1-0.6B$ is invertible and $\mathrm{Var}(\epsilon_t+\epsilon_t)=\mathrm{Var}(2\epsilon_t)$, the value of information sharing is zero even though $1-0.5B\neq1-0.7B$.

[17] Specifically, consider $X_t^1=(1-B+B^3)(1-0.5B)\epsilon_t$ and $X_t^2=(1-B+B^3)(1+0.5B)\delta_t$, where ϵ_t and δ_t are independent, and $\mathrm{Var}(\epsilon_t)=\mathrm{Var}(\delta_t)$. The aggregate process becomes $S_t=(1-B+B^3)\eta_t$, where $\eta_t=(1-0.5B)\epsilon_t+(1+0.5B)\delta_t$. Since $\tilde{\theta}_1=0$, the two-step-ahead forecast $\hat{S}_{t,t+2}$ is $-(1-0.5B)\epsilon_t-(1+0.5B)\delta_t+(1-0.5B)\epsilon_{t-1}+(1+0.5B)\delta_{t-1}+(1-0.5B)\epsilon_{t-2}+(1+0.5B)\delta_{t-2}$ for both cases.

[18] Recall that $\psi(B)=1+\gamma\kappa(B)(\alpha_0+\alpha_1B+\alpha_2B^2+\alpha_3B^3)$. X_t^1 and X_t^2 have a common factor if and only if $\varphi(B)=\kappa(B)(1-B)^{-1}$. If this is the case, the degree of $\psi(B)$ is larger than the degree of $(1-B)\pi(B)$, suggesting that the coefficient of the aggregate process is not their common factor. Thus, Assumption A2 holds.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

hold, and we can apply Theorem 2 to the case where the forecast lead time $h = 1$.

When $q_\epsilon = q_\delta = 0$, both X_t^1 and X_t^2 are i.i.d. processes, and thus, S_t is also an i.i.d. process. Any forecast is a constant, and thus there is no value from sharing the downstream sales information. This situation can only occur when $\varphi(B) = \pi(B) = \psi(B) = \kappa(B) = 1$, which means the retailer faces an i.i.d. demand and employs a demand replacement policy. In the rest of the paper, we will exclude this situation from the discussion, because, with i.i.d. demand, using historical observations cannot reduce any uncertainty of the future forecast.

If not both processes are i.i.d. models, or equivalently, if $q_\epsilon = q_\delta = 0$ is not true, then the two sets of parameters $c_0^{-1}\varphi(B)\psi(B)$ and $\pi(B)\kappa(B)$ can never be the same. The key ingredient in the proof is to show that the polynomial $(1 - B)$ is a factor in $\pi(B)\kappa(B)$ but not a factor in $c_0^{-1}\varphi(B)\psi(B)$, which leads to a positive value of information for any forecast lead time.

Compared with the invertibility conditions posted on the policy parameter $\psi(B)$ that induces positive value of information sharing in Proposition 1, Proposition 3 establishes a qualitatively different conclusion and intuition. Under a strict adherence to the inventory policy, the planner places orders based on the same information set that statisticians observe, which leads to a classical demand signal propagation studied in the literature. Our interview with the planner and our data suggests decision departures from the ideal policy, because retailers observe private information that is not observed by statisticians. Thus, unlike before, the demand now propagates together with decision deviations. Besides demand information, all decision deviation information should also be carried by the order process to produce a zero value of information sharing. Decision deviations, however, turn out to distort the normal demand propagation. The different propagation patterns of the demand process and decision deviation process drive the loss of information as they propagate upstream. To be specific, the ending inventory level carries the current week's decision deviation and rolls it over to the next week's replenishment decision that further determines the next week's ending inventory. Thus, the evolution of inventory governs the translation of exogenous decision deviation signals into orders. Demand signals, on the other hand, are governed by the evolution of both inventory and current demand. As both signals propagate together to become orders in such innately different patterns, the detailed information of the two processes is lost and is replaced with the less informative (larger uncertainty) order signals. Consequently, information sharing becomes valuable to recover the order's elaborate information structure and to forecast more accurately.

We conclude that when decision deviations and demand signals are both present, demand information is lost during propagation and orders are not a sufficient statistics of demand. Thus, it is impossible to infer demand from orders, or equivalently, demand cannot be written as a linear combination of *historical* orders. (Gaur et al. 2005 and Giloni et al. 2014 define this property as inferability.)

Generalization of Proposition 3. Our model assumes that the retailer has a weighted moving average demand forecast. We prove the same theoretical result when the demand forecast is optimal (shown in the technical companion). Furthermore, we prove that the value of information sharing is strictly positive under a more general setting where the retailer faces the MMFE demand and the GOUTP (shown in the technical companion). Most time-series demand models can be interpreted as a special case of the MMFE demand, such as the AR model, IMA model, the general ARMA model, and the linear state-space demand model (Chen and Lee 2009). The GOUTP covers a class of stationary and affine order-up-to policies, i.e., the myopic order-up-to policy, the production smoothing policy (Graves et al. 1998), and the ConDOI policy with ordering smoothing. It is worth noting that for any demand structure and replenishment policy that are an affine time-invariant combination of historical observations and signals, the key intuition that drives the positive value of information sharing remains: the evolution patterns of demand signals and decision deviation signals follow innately different evolution patterns. The distinct propagation patterns obscure the detailed information structure, which leads to a strictly positive value of information sharing for any linear demand model and inventory policy. We further strengthen the key result under more general demand and ordering policy settings.

Proposition 3 illustrates the value of information sharing when both demand and decision uncertainties are nonzero. If there is no decision deviation, Proposition 1 demonstrates the sufficient and necessary condition of positive value of information sharing. The following proposition, on the other hand, considers the other extreme case where demand uncertainty is zero.

Proposition 4. *When the demand shock is zero, the value of information sharing is zero for the one-step-ahead forecast.*

When demand is deterministic, the order becomes $O_t = m + \kappa(B)\delta_t$, where m is a constant composed of historical demand. Since only one signal series propagates to become orders, we can directly apply Proposition 1 to examine whether $\kappa(B)$ is invertible. The unique root of $\kappa(z^*) = 0$ lies on the unit circle, which

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

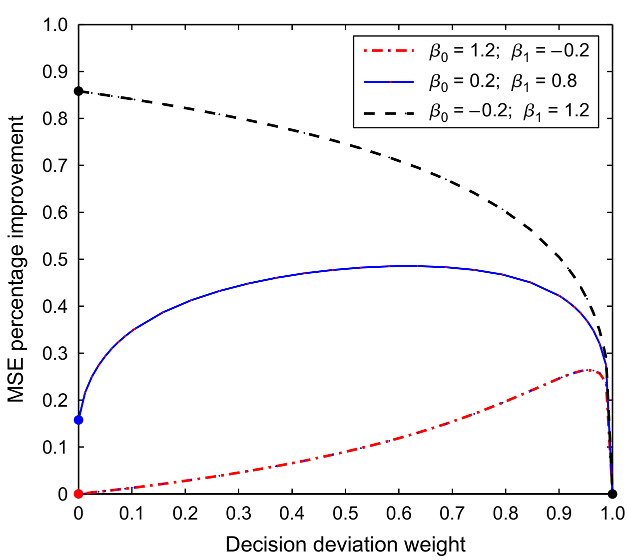
Plosser and Schwert (1997) defined as strictly noninvertibility. The author shows that the univariate MA parameter's estimator is asymptotically similar to the invertible processes, indicating that $\kappa(B)$ can be correctly estimated from historical orders. Therefore, we can infer the decision deviation history by orders and conclude zero value of information sharing.

To summarize the above theoretical findings, we characterize the value of information sharing with a numerical analysis. We measure the value of information sharing by the MSE percentage improvement of the InfoSharing forecasts over the NoInfoSharing forecasts. When $h = 1$, the measure is $(\mathrm{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O) - \mathrm{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O \cup \Omega_t^D))/\mathrm{Var}(O_{t+1} - \hat{O}_{t,t+1} \mid \Omega_t^O)$, which takes value between 0 and 1.

Figure 2 displays the MSE percentage improvement of the one-step-ahead forecast with respect to the relative weight of the decision deviation under three sets of inventory policy parameters. Keeping the DOI level and the order smoothing level fixed, we choose three sets of policy parameters β_i $(i = 0, 1)$ that correspond to three lines in Figure 2. Consider the retailer faces an ARIMA(0, 1, 1) demand, $D_t = D_{t-1} + \epsilon_t - \lambda\epsilon_{t-1}$, with the MA parameter $\lambda = 0.5$. The policy parameter $\psi(B)$ is noninvertible for the top two processes but invertible for the bottom one. In this numerical example, ϵ_t and δ_t are independent.

Our theoretical prediction aligns with the numerical observations. When the decision uncertainty is zero, the value of information sharing is positive for the first two and zero for the last policy parameters. This pattern is consistent with Proposition 1. Note that the studies in the literature correspond to the points on the vertical axis where the decision deviation weight is zero. Our theory can describe the entire curve. As decision deviations become dominant, there is little (and zero if decision deviation weight = 1) gain from sharing the downstream sales information, which coincides with Proposition 4. When both the decision uncertainty and the demand uncertainty exist, Figure 2 presents a strictly positive value of information sharing, which agrees with Proposition 3.

Figure 2 (Color online) MSE Percentage Improvement Against the Decision Deviation Weight for an ARIMA(0, 1, 1) Demand with $\lambda = 0.5$ and a ConDOI Policy with Order Smoothing with $\gamma = 0.8$ and $\Gamma = 2$



6.4. Our Theory is Supported by Our Empirical Findings

In this section, we validate whether the observed forecast accuracy improvements in the data set agrees with our theory predictions. To this end, we compare the predicted and actual root mean squared forecast error, separately under the InfoSharing method and the NoInfoSharing method.

To be specific, we first derive the actual value from the empirical study. Second, we calculate the predicted root mean squared forecast error based on our theory and the estimated demand and policy parameters. When there is information sharing, the mean squared forecast error is $c_0^2\sigma_\epsilon^2 + \sigma_\delta^2$. In the absence of information sharing, based on the innovation algorithm in the time-series literature, the mean squared forecast error becomes a function of c_0, σ_ϵ, σ_δ, policy parameters and demand parameters listed in Table 2.

We present the results in Figure 3. We plot the predicted against the actual root mean squared prediction error under both no information sharing (left) and information sharing (right) case. Each point represents a product with a corresponding index in Table 3. A perfect model fit would lead to the points lying on the 45-degree dashed line in the figure. We fit a regression of the theoretical predictions on the actual observations. Under the information sharing setting, predicted = −13.03 + 1.17 × actual, and the 95% confidence interval on the coefficient 1.17 is [0.95, 1.40]. Under the no information sharing setting, predicted = −14.02 + 1.03 × actual, and the 95% confidence interval on the coefficient 1.03 is [0.77, 1.29]. The predicted points from our model are overall close to the 45-degree line for both cases, indicating a good fit. It shows that our new theoretical model can well explain how demands propagate upstream, and thus can predict the value of information sharing well.

We have proved that the value of information sharing is strictly positive under any forecast lead time. In the following section, we study how its magnitude changes relative to key variables.

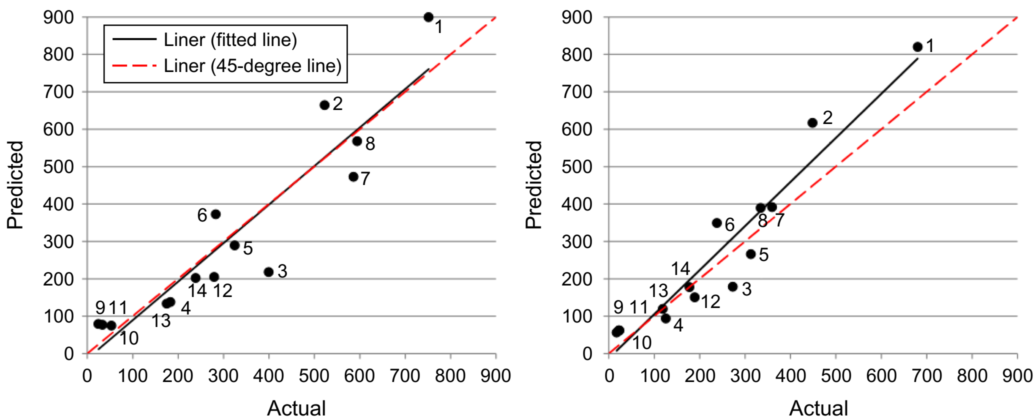
7. Theoretical Properties of the Value of Information Sharing

We present the magnitude of the benefit of information sharing as a function of the demand process characteristics, namely, λ of an ARIMA(0, 1, 1) demand

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Figure 3 (Color online) Consistency Between the Actual and Predicted Root Mean Squared Forecast Error Without Information Sharing (Left) and With Information Sharing (Right)



process (Lee et al. 2000 studies the impact of ρ in an AR(1) process). We then discuss the impact of other important variables, such as the forecast lead time and the order smoothing level. We focus on the one-step-ahead forecast. We theoretically analyze two special cases and resort to numerical studies for more involved settings.

We analyze a simple yet reasonable model to derive the theoretical prediction. The empirical estimation suggests that eight out of 14 products follow an ARIMA(0, 1, 1) demand. Henceforth, in this section, we focus on an ARIMA(0, 1, 1) demand with $\lambda \in [0, 1)$,

$$D_t = D_{t-1} + \epsilon_t - \lambda\epsilon_{t-1},$$

which can be equivalently written as an exponential smoothing form, $D_t = (1-\lambda)\sum_{i=1}^{\infty} \lambda^{i-1} D_{t-i} + \epsilon_t$. The current observation is a weighted average of historical observations with exponentially decaying coefficients. Values of λ closer to one put greater weight on recent data, and thus react more intensely to recent variations, whereas processes with λ closer to zero smooth the weight on past observations, and thus are less responsive to recent changes. Therefore, the process trends more slowly with a smaller λ. For example, the products that we study can be classified according to λ. Orange juice is an everyday drink for consumers. Sports drinks, on the other hand, are mainly consumed for exercising, and thus their consumptions are influenced by weather, temperature, and sporting events. The data exhibit a clearer slowly trending pattern in demand for sports drinks. Consistent with the above analysis, our estimated λ is larger for orange juice products and smaller for sports drinks products, according to our demand parameter estimations. We refer to demand with small λ as slowly trending demand.

Recall that the retailer's future demand forecast is a weighted average of historical $H+1$ weeks' demands. In the rest of this section, we assume the retailer's order relies on current and last weeks' demand, $H = 1$. We assume that demand signals and decision deviations are independent. The order can be written as a summation of two processes in ϵ_t and δ_s as in (12),

$$X_t^1 = (1 + \gamma\Gamma\beta_0 + \gamma\Gamma\beta_1 B)(1+\lambda B)\epsilon_t - \gamma^2 \sum_{i=1}^{\infty} (1-\gamma)^{i-1} (\Gamma\beta_0 + \Gamma\beta_1 B)(1+\lambda B)\epsilon_{t-i}$$

and

$$X_t^2 = \delta_t - \delta_{t-1} - \sum_{i=1}^{\infty} \gamma(1-\gamma)^{i-1}(\delta_{t-i} - \delta_{t-i-1}).$$

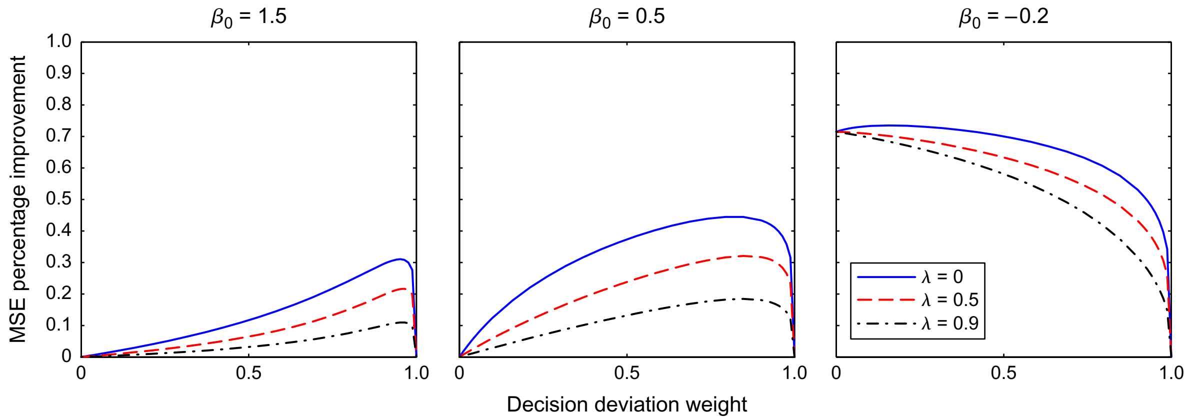
We focus on processes X_t^1 and X_t^2 with degrees smaller or equal to 3. When the degree of either process exceeds 3, the complexity of the problem precludes analytically tractable solutions and necessitates numerical analysis. Note that we will relax this restriction by allowing an infinite degree in the following numerical study. Therefore, we study two simple policies: (1) the retailer follows a demand replacement policy ($\gamma = 0$), and (2) the retailer adopts a ConDOI policy ($\gamma = 1$) with zero weight on the previous week's demand ($\beta_1 = 0$). Under (1), the order process is $O_t = D_t + \delta_t$, and under (2), the order process is $O_t = (1 + \Gamma\beta_0)D_t - \Gamma\beta_0 D_{t-1} + \delta_t - \delta_{t-1}$. The following proposition demonstrates that the value strictly decreases with λ (see the proof in the technical companion).

Proposition 5. *The value of information sharing under the one-step-ahead forecast strictly decreases with λ under both* (1) *and* (2).

To further explore the demand's impact under other policy parameters, we conduct numerical studies. Figure 4 presents the relation of MSE percentage improvement with respect to λ under three policy weight parameters ($\beta_0 = 1.5$, 0.5, and -0.2). The value

RIGHTSLINK

Figure 4 (Color online) Under an ARIMA(0, 1, 1) Demand with λ and a ConDOI Policy with Order Smoothing with β_0, $\gamma = 0.5$ and $\Gamma = 2$, the MSE Percentage Improvement Strictly Increases with λ



declines as λ is larger, which corroborates the analytical findings without restrictions on the degree of two MA processes. When λ is closer to one, the coefficient of demand signals becomes closer to the coefficient of decision deviations, and hence, less information is lost as signals propagate upstream and information sharing is less valuable. For example, under the demand replacement policy (1), when $\lambda = 1$, the centered order becomes $\epsilon_t - \epsilon_{t-1} + \delta_t - \delta_{t-1}$. An immediate result from Theorem 2 is a complete information propagation and zero value of demand.

Let us revisit the empirical MSE percentage improvement in the last column of Table 3. Except for the two orange drinks 12 OR and 12 ORCA, which have a much smaller bottle volume compared to other orange juice products and serve a similar function as sports drinks, the orange juice products gain less from information sharing than the rest of the products. Consistent with our theoretical predictions, their λ is closer to zero, which differs substantially from 12 OR, 12 ORCA, and the other sports drinks in Table 2. Hence, our theory can provide a correct mapping from the demand pattern to the potential gain from information sharing.

The result implies that it is more worthwhile for suppliers in industries with slow trending consumptions to invest in the information sharing system to improve predictions. Note that forecasting beyond one week might reverse the relation of the value of information sharing and demand parameter λ. We recommend that managers resort to run a numerical study to validate the potential gain based on demand and policy characteristics.

Forecast Lead Time. Note that the h-step-ahead forecast is the sum of forecasts within h periods. Companies may also monitor predictions in a specific lead time, which they use to adjust manufacture plans. We define the hth-step-ahead forecast as the forecast made in period t about demand in period $t + h$. We conduct a numerical analysis on the impact of the forecast lead time (see detailed results in the technical companion), and we find different results on the two forecast metrics. We show that the value of information sharing of the hth-step-ahead forecast strictly decreases as the forecast lead time increases. This is because future signals are less dependent on historical demand, and thus, the future uncertainty is less likely to be resolved with information sharing. This implies a limited potential gain in farther forecasts. We find that for the h-step-ahead forecast, the value of information sharing might increase with forecast lead time under some conditions. This is because historical signals are cancelled out when summing the future forecasts within h periods, and the forecast error might increase less slowly with forecast lead time under information sharing.[19]

Order Smoothing Level. We also explore the impact of the smoothing level (see numerical studies in the technical companion). We find that the value of information sharing increases as γ increases. When γ is close to one, there is less inventory smoothing, and the order depends more on historical demand and inventory. Thus, current demand becomes very valuable for predicting future orders, and information sharing provides greater improvements. It is interesting to note that, when $\gamma = 0$ (demand replacement policy with $O_t = D_t + \delta_t$), the order only relies on demand in the period, which leads to a small benefit from sharing information.

[19] For example, when the decision deviation process is $\delta_t - \delta_{t-1}$, its h-step-ahead forecast error is δ_{t+h} and the variance does not change with forecast lead time. In contrast, the aggregate process does not preserve this structure, and the h-step-ahead forecast error might strictly increases with forecast lead time. If the decision deviation weight is relatively large, the forecast error under information sharing hardly increases with forecast lead time, which means that farther forecast can have a higher value of information.

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

8. The Value of Operational Knowledge

Note that in previous sections, apart from incorporating demand information, we also explicitly used knowledge of the downstream replenishment policy. We define it as operational knowledge. In our study, operational knowledge is ConDOI policy with order smoothing where the retailer's demand forecast is a weighted moving average of recent four weeks' demand. When the supplier does not have such knowledge, the value of demand might be limited. (Chen and Lee 2009 propose that the retailer should share projected future orders to avoid sharing the policy.) In this section, we empirically quantify the benefit of knowing the downstream replenishment policy, which we refer to as operational knowledge. We replicate the situation where companies lack operational knowledge, using statistical forecasting methods that capture statistical correlations between sales and orders. By contrasting them against the InfoSharing and NoInfoSharing methods, we decompose the value of information sharing into two parts: the value stemming from only sales information and the additional value stemming from operational knowledge. The NoInfoSharing forecast only uses order information. Its forecast accuracy difference from the statistical forecasts measures the benefit of only sales data. The InfoSharing forecast incorporates downstream sales information as well as the downstream inventory policy. Its forecast accuracy improvement over the statistical forecasts measures the value of operational knowledge.

We consider three forecasting methods. We first capture the order demand correlation by regressing orders on demands within the past five weeks, referred to as the Reg *D* method:

$$O_t = c_0 D_t + c_1 D_{t-1} + \cdots + c_5 D_{t-5} + \varepsilon_t. \tag{13}$$

Similarly, we regress orders on both orders and demands, and we call it the Reg *D* and *O* method:

$$O_t = c_0 D_t + c_1 D_{t-1} + \cdots + c_5 D_{t-5} + b_1 O_{t-1} + \cdots + b_5 O_{t-5} + \varepsilon_t. \tag{14}$$

We then treat the order as an ARIMA process while at the same time accounting for demand, which adds observed demands to orders in (10):

$$\begin{aligned}(1-B)^d O_t &= \mu + \tilde{\rho}_1 (1-B)^d O_{t-1} + \cdots + \tilde{\rho}_p (1-B)^d O^1_{t-p} \\ &\quad + \eta_t + \tilde{\lambda}_1 \eta_{t-1} + \cdots + \tilde{\lambda}_q \eta_{t-q} + c_1 (1-B)^d D_t \\ &\quad + \cdots + c_p (1-B)^d D_{t-p+1} + \epsilon_{t+1} \\ &\quad + e_1 \epsilon_t + \cdots + e_q \epsilon_{t-q+1}.\end{aligned} \tag{15}$$

The parameters in Equation (15) can be estimated by fitting O_t and D_{t+1} series in a two-dimensional vector ARIMA model. Note that method (15) is more general than method (10). We specify a vector ARIMA$(3,1,1)$[20] model with $\mu = 0$. We refer to this as the vector ARIMA method.

Table 4 Forecast Accuracy Summary for All Methods at the Overall Level

	NoInfo Sharing	Vector ARIMA	Reg *D*	Reg *D* and *O*	Info Sharing
MAPE (%)	56.45	42.72*	45.94*	42.18*	**33.36***

Notes. Significant accuracy improvement over the NoInfoSharing method is marked by an asterisk ($p = 0.01$). Significant ($p = 0.05$) accuracy improvement of the InfoSharing method over the other unbold methods is also in bold. All methods that include downstream sales outperform the NoInfoSharing forecast. The value of operational knowledge is positive, since the InfoSharing forecast outperforms any statistical method.

Let us summarize the overall prediction accuracy in MAPE across all products in Table 4. The NoInfoSharing forecasts have the lowest accuracy (or highest forecast error) with 56.45% MAPE, and the InfoSharing forecasts have the highest predictive power with 33.36% MAPE. We can achieve a 40.90% percentage improvement in total. We show that 21.26% comes from statistical methods and 19.64% comes from operational knowledge. This means that operational knowledge, as least in our study, brings a similar order of magnitude of forecasting improvements as using only sales information. This suggests that companies should always learn the downstream replenishment policy to specify the correct structure between orders and demand, which enables the companies to better utilize downstream demand information to achieve the greatest improvements.

Remark. One might notice that the only difference between Equation (9) and Equation (13) is I_{t-1}. This does not mean that operational knowledge is only about knowing or including retail inventory. Operational knowledge is about using the policy structure to incorporate retail inventory in the order decisions. Note that we quantify the value of operational knowledge in a specific setting. It is possible that operational knowledge can bring no value in some cases. For

[20] VARIMA$(3,1,1)$ model is

$$\begin{bmatrix} O^d_t \\ D^d_{t+1} \end{bmatrix} = \begin{bmatrix} c^1_{11} & c^1_{12} \\ c^1_{21} & c^1_{22} \end{bmatrix} \begin{bmatrix} O^d_{t-1} \\ D^d_t \end{bmatrix} + \cdots + \begin{bmatrix} c^3_{11} & c^3_{12} \\ c^3_{21} & c^3_{22} \end{bmatrix} \begin{bmatrix} O^d_{t-3} \\ D^d_{t-2} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_{t+1} \end{bmatrix} + \begin{bmatrix} e^1_{11} & e^1_{12} \\ e^1_{21} & e^1_{22} \end{bmatrix} \begin{bmatrix} \eta_{t-1} \\ \epsilon_t \end{bmatrix},$$

where c^i_{21} and c^i_{22} are restricted to zero for $i = 1, 2, 3$; e^1_{12}, e^1_{21}, and e^1_{22} are restricted to zero; η_t is the order shock; and ϵ_t is the demand shock. We choose $(3,1,1)$ because of the computational constraints. Such parameters can represent the majority of parameters found in (10).

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

example, when the retailer follows a demand replacement policy, we have $O_t = D_t + \varepsilon_t$. Simply estimating the correlation between O_t and D_t can correctly capture the policy parameters, and there is no value of operational knowledge.

9. Conclusions and Discussions

This paper empirically evaluates the supplier's forecast improvement by incorporating downstream retail sales data and supports the observations with an extended theoretical model. Table 3 in §5.2 summarizes our main empirical findings. Our observations highlight the positive value to suppliers of incorporating retailers' sales data: 4.7%–58.6% MAPE percentage improvements across 14 products and a 40.90% improvement in MAPE on an overall level, which is regarded as a significant improvement by the CPG company we studied. In addition, we empirically decompose the 40.90% total improvement into two parts, 21.26% from sales data and 19.64% from knowledge of the downstream replenishment policy.

We revisit and extend the theoretical model in the existing literature. Until now, the theoretical literature showed no value of information sharing for 10 out of 14 products. By taking into account the idiosyncratic shocks in decision making, we relax the strict adherence to the replenishment policy assumption in the literature. Our new theory yields qualitatively different results than the previous literature. We demonstrate that if both demand signals and decision signals are nonzero, the value of information sharing is strictly positive for any forecast lead time. Further, our new theory is supported by our empirical observations. Our paper, therefore, underscores the importance of recognizing that the decision maker may deviate from the exact policy, a phenomenon that is common in practice and is absent in earlier theoretical models.

We show that decision deviations are present and have a significant impact on the value of information sharing. We identified several possible operational explanations for the deviations. Interesting studies could also look at behavioral factors that would cause decision deviations (see studies by Schweitzer and Cachon 2000 and Van Donselaar et al. 2010).

Our paper suggests that the value of information sharing is higher when (1) product demand has high local volatility, such as orange juice that is consumed daily; (2) the retailer's ordering policy has a low inventory smoothing level, such as a strict ConDOI policy; or (3) incorporating the inventory policy structure in determining the relationship between orders and demand.

Our study focuses on a specific linear and stationary inventory policy with a stationary demand process. It is worth noting that the conclusion regarding the strictly positive value of information sharing can be generalized to any linear and stationary inventory policy and stationary demand. In fact, we prove that the value of information sharing is always positive, if a retailer follows the generalized order-up-to policy and the MMFE demand (in the technical companion). We highlight the key intuition that continues to remain in general settings: demand signals and decision deviations accumulate in innately different evolution patterns as they propagate upstream.

In this paper, we focus on low-promotional products, the demand and orders of which are stationary. We also empirically test the forecast accuracy of the remaining high-promotional products (we elaborate on the forecasting procedure and results in the technical companion). We find that the value of information sharing is positive for most promotional products. As the promotional depth increases, the forecast accuracy decreases for both scenarios (with and without information sharing), because the order series has higher uncertainty. However, we observe an insignificant correlation between forecast improvements and the promotional depth. A potential reason comes from the nonstationary order structure (in demand signals and decision deviation signals) caused by nonstationary price promotions. The optimal estimators (of the ARIMA model for orders or the replenishment policy) obtained in the current week might be suboptimal for the future, which might affect the forecast precision of the two scenarios differently. Another potential reason is that, with information sharing, the estimating equation of the replenishment policy might not correctly estimate parameters for the high promotional products, because the method by which the retailer forecasts future demand and how it determines orders becomes more complicated than the policy assumed in our model. Future research is needed to understand how promotional activities affect the information transmission and the value of sharing downstream demand.

Our model can represent many industries in practice. Our results are applicable to other types of products when their demand and replenishment policy follows a linear structure. However, our analysis has limitations. In particular, future research should break the affine structure and explore nonlinear policies such as the (s, S) policy. This requires a reexamination via a nonlinear time-series model or a proper approximation.

For practitioners, having access to more data points, such as the daily level data, can help identify detailed ordering patterns and benefit the forecast precision. For example, one can distinguish the "Friday effect" and the "batching effect" in a positive decision deviation (overreplenishment from the target inventory policy) by the day of the order. However, forecasting

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

at the daily level comes at the price of both a heavy computation burden and a nonlinear estimation. The CPG company currently needs around 24 hours on modern computing technology starting each Saturday night to run the ARIMA model on all product lines. If the company were to forecast at the daily level, they would need more computational power or would need to start their forecasts earlier (than Saturday night), which would increase their forecast lead time, and thus reduce their forecast accuracy. In addition, the daily level data has a lot of zero orders (the planner claims to order only once per week around 80% of the weeks). For the baseline forecast without downstream demand, fitting linear time-series models is not a good option on such a data set. Further, to obtain the forecast with downstream demand, we need to model a nonlinear inventory policy, which takes more effort to estimate. Nonlinear models are beyond the scope of our paper and require further investigation. Forecasting at a coarse level, such as the monthly level or beyond, loses precision and thus fails to support short-term managerial decisions.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.2132.

Appendix

Proof of Theorem 2. *With information sharing.* Recall that $X_t^i=\chi_i(B)\epsilon_t^i$ with coefficient $\chi_i(B)=1+\lambda_1^iB+\lambda_2^iB^2+\cdots+\lambda_{q_i}^iB^{q_i}$. Let $\Omega_t^{X^i}=\operatorname{span}\{\epsilon_1^i,\ldots,\epsilon_t^i\}$ denote the plane containing the historical shocks $\epsilon_1^i,\ldots,\epsilon_t^i$. According to the definition, $\epsilon_{t+l}^i\perp\Omega_t^{X_i}$ for any $l\geq 1$. Since we assume $\epsilon_t^i\perp\epsilon_{t-k}^j$ for any $k>0$, the general orthogonal condition can be written as

$$\epsilon_{t+l}^i\perp\Omega_t^{X_j},\quad \forall\, i,j,l. \tag{16}$$

The h-step-ahead forecast of process X_t^i made in period t is $\hat{X}_{t,t+h}^i=\lambda_h^i\epsilon_t+\lambda_{h+1}^i\epsilon_{t-1}^i+\cdots+\lambda_{q_i}^i\epsilon_{t+h-q_i}^i$. The total forecast with information sharing is $\sum_{i=1}^N\hat{X}_{t,t+h}^i$. We denote

$$\operatorname{Var}\bigg(\sum_{l=1}^h(S_{t+l}-\hat{S}_{t,t+l})\,\Big|\,\bigcup_i\Omega_t^{X^i}\bigg)\quad\text{as}\quad\operatorname{Var}\bigg(\sum_{l=1}^h\Big(S_{t+l}-\sum_{i=1}^N\hat{X}_{t,t+l}^i\Big)\bigg)$$

in the proof.

Without information sharing. In absence of demand information, the order process is $S_t=\chi_S(B)\eta_t$, where $\chi_S(B)=1+\theta_1B+\theta_2B^2+\cdots+\theta_{q_S}B^{q_S}$. (Note that q_S might be infinite.) Let $\Omega_t^S=\operatorname{span}\{\eta_1,\ldots,\eta_t\}$ denote the plane containing the order process signals $\eta_1,\ldots,\eta_t$. Since $\chi_S(B)\eta_t=\sum_{i=1}^N\chi_i(B)\epsilon_t^i$ and $\chi_S(B)$ is invertible, then $\eta_t=\chi_S^{-1}(B)\sum_{i=1}^N\chi_i(B)\epsilon_t^i$. We have $\Omega_t^S\in\bigcup_i\Omega_t^{X^i}$, because η_t is a linear combination of ϵ_s^i, $s\leq t$.

According to (16), we have $\epsilon_{t+l}^j\perp\bigcup_i\Omega_t^{X^i}$ for any j and any $l\geq 1$. Since $\Omega_t^S\in\bigcup_i\Omega_t^{X^i}$, then

$$\epsilon_{t+l}^i\perp\Omega_t^S,\quad\forall\, i,l. \tag{17}$$

The h-step-ahead forecast of process S_t made in period t is $\hat{S}_{t,t+h}=\theta_h\eta_t+\theta_{h+1}\eta_{t-1}+\cdots+\theta_{q_S}\eta_{t+h-q_S}$. We denote $\operatorname{Var}(\sum_{l=1}^h(S_{t+l}-\hat{S}_{t,t+l})\,|\,\Omega_t^S)$ as $\operatorname{Var}(\sum_{l=1}^h(S_{t+l}-\hat{S}_{t,t+l}))$ in the proof.

The Value of Information Sharing. We first prove the statement under the assumption $\epsilon_t^i\perp\epsilon_t^j$ for any $i\neq j$. We rewrite $\operatorname{Var}(S_{t+h}-\hat{S}_{t,t+h})$ as $\operatorname{Var}(S_{t+h}-\sum_{i=1}^N\hat{X}_{t,t+h}^i+\sum_{i=1}^N\hat{X}_{t,t+h}^i-\hat{S}_{t,t+h})$. According to the orthogonal condition (16) and (17), $\operatorname{Var}(\sum_{l=1}^h(S_{t+l}-\hat{S}_{t,t+l}))$ can be simplified to

$$\operatorname{Var}\bigg(\sum_{l=1}^h(S_{t+l}-\hat{S}_{t,t+l})\bigg)=\operatorname{Var}\bigg(\sum_{l=1}^h\Big(S_{t+l}-\sum_{i=1}^N\hat{X}_{t,t+l}^i\Big)\bigg)+\operatorname{Var}\bigg(\sum_{l=1}^h\Big(\hat{S}_{t,t+l}-\sum_{i=1}^N\hat{X}_{t,t+l}^i\Big)\bigg). \tag{18}$$

$\operatorname{Var}(\sum_{l=1}^h(S_{t+l}-\hat{S}_{t,t+l}))>\operatorname{Var}(\sum_{l=1}^h(S_{t+l}-\sum_{i=1}^N\hat{X}_{t,t+l}^i))$ if and only if $\sum_{l=1}^h(\hat{S}_{t,t+l}-\sum_{i=1}^N\hat{X}_{t,t+l}^i)\neq 0$. We prove that if $\chi_i(B)\neq\chi_j(B)$ for some i,j, then $\sum_{l=1}^h(\hat{S}_{t,t+l}-\sum_{i=1}^N\hat{X}_{t,t+l}^i)\neq 0$.

Suppose that there exists a finite forecast lead time h, where $h\leq\max_i\{q_i\}$ such that

$$\sum_{l=1}^h\hat{S}_{t,t+l}=\sum_{l=1}^h\sum_{i=1}^N\hat{X}_{t,t+l}^i.$$

This is equivalent to $\sum_{l=1}^h\hat{S}_{t-h,t+l-h}=\sum_{l=1}^h\sum_{i=1}^N\hat{X}_{t-h,t+l-h}^i$, which can be expanded as

$$\begin{aligned}&\bigg(\sum_{j=1}^h\theta_j\bigg)\eta_{t-h}+\bigg(\sum_{j=2}^{h+1}\theta_j\bigg)\eta_{t-h-1}+\cdots+(\theta_{q_S-1}+\theta_{q_S})\eta_{t-q_S-h+2}\\&\quad+\theta_{q_S}\eta_{t-q_S-h+1}\\&\quad=\sum_{i=1}^N\bigg(\bigg(\sum_{j=1}^h\lambda_j^i\bigg)\epsilon_{t-h}+\bigg(\sum_{j=2}^{h+1}\lambda_j^i\bigg)\epsilon_{t-h-1}^i+\cdots\\&\qquad+(\lambda_{q_i-1}^i+\lambda_{q_i}^i)\epsilon_{t-q_i-h+2}^i+\lambda_{q_i}^i\epsilon_{t-q_i-h+1}^i\bigg).\end{aligned}$$

For notational convenience, let $\lambda_j^i=0$ for $j>q_i$. We define $\sum_{j=0}^k\theta_j$ as $\tilde{\theta}_k$ and $\sum_{j=0}^k\lambda_j^i$ as $\tilde{\lambda}_k^i$. Given $S_t=\sum_{i=1}^NX_t^i$, we subtract the above equation from $\sum_{r=t}^{t-h+1}S_r=\sum_{r=t}^{t-h+1}\sum_{i=1}^NX_r^i$:

$$\begin{aligned}&\eta_t+\tilde{\theta}_1\eta_{t-1}+\cdots+\tilde{\theta}_{h-1}\eta_{t-h+1}\\&\quad=\sum_{i=1}^N(\epsilon_t^i+\tilde{\lambda}_1^i\epsilon_{t-1}^i+\cdots+\tilde{\lambda}_{h-1}^i\epsilon_{t-h+1}^i).\end{aligned}\tag{19}$$

Since $\chi_S(B)$ is invertible, $\chi_S^{-1}(B)$ has finite degree. We replace η_{t-j} with $\chi_S^{-1}(B)\sum_{i=1}^N\chi_i(B)\epsilon_{t-j}^i$ for all j:

$$\eta_t+\tilde{\theta}_1\eta_{t-1}+\cdots+\tilde{\theta}_{h-1}\eta_{t-h+1}=\sum_{j=0}^{h-1}\tilde{\theta}_j\chi_S^{-1}(B)\sum_{i=1}^N\chi_i(B)\epsilon_{t-j}^i. \tag{20}$$

According to Assumption A2, there exists at least one process k such that $\chi_S^{-1}(B)\chi_k(B)$ is of infinite degree. Therefore, the degree with respect to ϵ_t^k is infinite in (20) whereas the degree is finite in (19). We have reached a contradiction. Therefore, the value of information sharing is positive.

Note that in the above analysis, we use the assumption that the shocks are independent across processes. We next extend the above proof to the situation where contemporaneous signals are correlated. We prove that, when there exist two processes with different coefficients, S_t equals to

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 00:30 . For personal use only, all rights reserved.

the sum of N time-series processes with orthogonal signals and the two processes still have different coefficients. We denote the two different processes as X_t^{N-1} and X_t^N, where $\chi_{N-1}(B) \neq \chi_N(B)$. Since Assumption A1 requires that ϵ_t^i is not a linear combination of ϵ_t^{-i} for any i, we can decompose ϵ_t^2 into $\beta_{2,1}\epsilon_t^1$ and $\hat{\epsilon}_t^2$, where $\hat{\epsilon}_t^2 \perp \epsilon_t^1$. For the same reason, we can decompose ϵ_t^i into $\beta_{i,1}\epsilon_t^1$, $\beta_{i,2}\hat{\epsilon}_t^2, \ldots$ and $\hat{\epsilon}_t^i$, where $\hat{\epsilon}_t^i \perp \epsilon_t^j$ for any $j < i$. The aggregate process can be written as N processes with orthogonal signals, and the last two processes are $(\chi_{N-1}(B) + \beta_{N,N-1}\chi_N(B))\hat{\epsilon}_t^{N-1}$ and $\chi_N(B)\hat{\epsilon}_t^N$.

If $\beta_{N,1} \neq 0$, then $\chi_{N-1}(B) + \beta_{N,N-1}\chi_N(B) \neq (1+\beta_{N,1})\chi_N(B)$, and thus their coefficients are different. If $\beta_{N,N-1} = 0$, then according to the assumption, we have $\chi_{N-1}(B) \neq \chi_N(B)$. We can apply the above results to show that for $h \leq \max_i\{q_i\}$, the value is positive. □

References

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

Aviv Y (2003) A time-series framework for supply chain inventory management. *Oper. Res.* 51(2):210–227.

Balakrishnan A, Geunes J, Pangburn MS (2004) Coordinating supply chains by controlling upstream variability propagation. *Manufacturing Service Oper. Management* 6(2):163–183.

Bray RL, Mendelson H (2012) Information transmission and the bullwhip effect: An empirical investigation. *Management Sci.* 58(5):860–875.

Brockwell PJ, Davis RA (2002) *Introduction to Time Series and Forecasting* (Springer, New York).

Cachon GP, Fisher M (2000) Supply chain inventory management and the value of shared information. *Management Sci.* 46(8):1032–1048.

Cachon GP, Randall T, Schmidt GM (2007) In search of the bullwhip effect. *Manufacturing Service Oper. Management* 9(4): 457–479.

Caplin AS (1985) The variability of aggregate demand with (s, S) inventory policies. *Econometrica* 53(6):1395–1409.

Chen F, Ryan JK, Simchi-Levi D (2000a) The impact of exponential smoothing forecasts on the bullwhip effect. *Naval Res. Logist.* 47(4):269–286.

Chen F, Drezner Z, Ryan JK, Simchi-Levi D (2000b) Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Sci.* 46(3):436–443.

Chen L, Lee HL (2009) Information sharing and order variability control under a generalized demand model. *Management Sci.* 55(5):781–797.

Cohen MA, Ho TH, Ren ZJ, Terwiesch C (2003) Measuring imputed cost in the semiconductor equipment supply chain. *Management Sci.* 49(12):1653–1670.

Dong Y, Dresner M, Yao Y (2014) Beyond information sharing: An empirical analysis of vendor-managed inventory. *Production Oper. Management* 23(5):817–828.

Gaur V, Giloni A, Seshadri S (2005) Information sharing in a supply chain under ARMA demand. *Management Sci.* 51(6):961–969.

Gilbert K (2005) An ARIMA supply chain model. *Management Sci.* 51(2):305–310.

Giloni A, Hurvich C, Seshadri S (2014) Forecasting and information sharing in supply chains under ARMA demand. *IIE Trans.* 46(1):35–54.

Graves SC (1999) A single-item inventory model for a nonstationary demand process. *Manufacturing Service Oper. Management* 1(1):50–61.

Graves SC, Kletter DB, Hetzel WB (1998) A dynamic model for requirements planning with application to supply chain optimization. *Oper. Res.* 46(3):S35–S49.

Hamilton JD (1994) *Time Series Analysis* (Princeton University Press, Princeton, NJ).

Heath DC, Jackson PL (1994) Modelling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Trans* 26(3):17–30.

Irvine FO (1981) Retail inventory investment and the cost of capital. *Amer. Econom. Rev.* 71(4):633–648.

Kesavan S, Gaur V, Raman A (2010) Do inventory and gross margin data improve sales forecasts for U.S. public retailers? *Management Sci.* 56(9):1519–1533.

Ledesma G (2004) Waste not, want not. *Material Handling & Logistics* (April 1), http://mhlnews.com/global-supply-chain/waste-not-want-not.

Lee HL, Padmanabhan V, Whang S (1997) Information distortion in a supply chain: The bullwhip effect. *Management Sci.* 43(4): 546–558.

Lee HL, So KC, Tang CS (2000) The value of information sharing in a two-level supply chain. *Management Sci.* 46(5):626–643.

Miyaoka J, Hausman W (2004) How a base stock policy using "stale" forecasts provides supply chain benefits. *Manufacturing Service Oper. Management* 6(2):149–162.

Osadchiy N, Gaur V, Seshadri S (2013) Sales forecasting with financial indicators and experts' input. *Production Oper. Management* 22(5):1056–1076.

Plosser CI, Schwert GW (1997) Estimation of a non-invertible moving average process: The case of overdifferencing. *J. Econometrics* (6):199–244.

Retail Info Systems News (2013) Costco improves CRX supplier data technology platform. (April 22), http://risnews.edgl.com/retail-news/Costco-Improves-CRX-Supplier-Data-Technology-Platform86002.

Rust J (1994) Structural estimation of Markov decision processes. Engle RF, McFadden DL, eds. *Handbook of Econometrics*, Vol. 4 (North-Holland, Amsterdam), 3081–3143.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Terwiesch C, Ren ZJ, Ho TH, Cohen MA (2005) An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Sci.* 51(2):208–220.

Van Donselaar KH, Gaur V, Van Woensel T, Broekmeulen RACM, Fransoo JC (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Sci.* 56(5):766–784.

Wold H (1938) *A Study in the Analysis of Stationary Time Series* (Almqvist and Wiksell, Uppsala, Sweden).

RIGHTSLINK
Copyright Clearance Center