Technical Companion to:

Information Sharing in Supply Chains: An Empirical and Theoretical Valuation

Nov 15, 2014

This technical companion serves as a supplementary material. We first show the proof of Proposition 5 of the paper when the retailer follows a ConDOI policy. We extend our main theoretical conclusions to three settings: (1) the retailer follows a ConDOI policy with an optimal demand forecast; (2) the retailer follows the general linear replenishment policy (GOUTP) and general demand (MMFE); and (3) the product is subject to price promotions. We empirically measure the forecast accuracy of high-promotional products and display the forecast accuracy improvement summary for all methods at the product level. We then discuss two properties of the value of information: the impact of the replenishment policy and the forecast lead time. Finally, we show the proofs of propositions in the paper.

1 Theorem and proof of the paper

We next show that when Assumption $A3$ is violated, the result of Theorem 2 can still hold as long as we post an assumption on the forecast lead time. That is, when $\chi_S^{-1}(B)\chi_i(B)$ is of finite degree for any i, the value of information sharing is not positive for any forecast lead time. Let $\tilde{\theta}_k$ denote $\sum_{j=0}^{k}\theta_j$, and let q_{i-S} denote the degree of $\chi_S(B)^{-1}\chi_i(B)$. We define j as $arg\max_i q_i$. We define a set of forecast lead time H that satisfies $\tilde{\theta}_{h-n}=0$ for $1 \leq n \leq q_{j-S}$.

Theorem 1 *Under $A1$, when $A2$ is not satisfied, if there exist two processes with different coefficients, $\chi_i(B) \neq \chi_j(B)$ for some i, j, then* $\mathrm{Var}(\sum_{l=1}^{h}(S_{t+l}-\hat{S}_{t,t+l})|\cup_i \Omega_t^{X^i}) < \mathrm{Var}(\sum_{l=1}^{h}(S_{t+l}-\hat{S}_{t,t+l})|\Omega_t^S)$ *for any finite forecast lead time h, where $h \leq \max_i\{q_i\}$ and $h \notin H$.*

Proof: The proof of Theorem 2 (in the main body of the paper) remains the same, except for the following changes on the contradiction of the degree between (19) and (20). We now assume that $\chi_S^{-1}(B)\chi_i(B)$ is of finite degree for any i. Since there exists n such that $\tilde{\theta}_{h-n} \neq 0$ for $1 \leq n \leq q_{j-S}$, the degree with respect to ϵ_t^j in equation (20) of the paper is at least $q_{i-S}+h-q_{i-S}$, which is equivalent to h. According to (19), the degree with respect to ϵ_t^j is $h-1$. As a result, for any finite forecast lead time $h \leq \max_i\{q_i\}$, the degree with respect to ϵ_t^j is strictly larger in (20). We have reached a contradiction.■

Note that $h = 1$ does not satisfy H, and this means that when $A2$ is violated, the result of Theorem 2 holds for the 1-step-ahead forecast.

Proof of Proposition 5

The retailer follows the demand replacement policy. Recall that the centered order is the summation of two MA processes $O_t - O_{t-1} = \epsilon_t - \lambda\epsilon_{t-1} + \delta_t - \delta_{t-1}$. We denote the aggregate MA process (or the order process) as $S_t = \eta_t + \theta\eta_{t-1}$, where $\mathrm{Var}(\eta_t) = v$. It satisfies the covariance equations, $-\lambda\sigma_\epsilon^2 - \sigma_\delta^2 = v\theta$ and $(1+\lambda^2)\sigma_\epsilon^2 + 2\sigma_\delta^2 = v(1+\theta^2)$. We substitute θ with v in the above equation to obtain a function $f(v,\lambda)$ with variable v and parameter λ such that the solution of $f(v,\lambda) = 0$ is the variance of η_t. The function $f(v,\lambda)$ satisfies $f(v,\lambda) = v^2 - ((1+\lambda^2)\sigma_\epsilon^2 + 2\sigma_\delta^2)v + (-\lambda\sigma_\epsilon^2 - \sigma_\delta^2)^2$.

The aggregate process S_t has invertible and noninvertible representations. Fixing λ, $f(v,\lambda) = 0$ has two solutions: the variance of the invertible process and the variance of the noninvertible process. We denote the former as v^*. The value of information sharing is $1 - (\sigma_\epsilon^2 + \sigma_\delta^2)/v^*$. We will prove that v^* is decreasing in λ.

We have $\partial f(v,\lambda)/\partial v = 2v - ((1+\lambda^2)\sigma_\epsilon^2 + 2\sigma_\delta^2)$. Since the invertible solution θ is smaller than 1, $2v^* = 2((1+\lambda^2)\sigma_\epsilon^2 + 2\sigma_\delta^2)/(1+\theta^2) > (1+\lambda^2)\sigma_\epsilon^2 + 2\sigma_\delta^2$. Thus, $\partial f(v,\lambda)/\partial v > 0$ at v^*. Since $f(v,\lambda)$ is continuous, $\partial f(v,\lambda)/\partial v > 0$ in an open interval of $v^*(\lambda)$. v^* is decreasing in λ. Therefore, it suffices to show that $\partial f(v,\lambda)/\partial\lambda > 0$. We have

$$\frac{\partial f(v,\lambda)}{\partial\lambda} = 2\sigma_\epsilon^2(-\lambda v + \lambda\sigma_\epsilon^2 + \sigma_\delta^2) = -2\sigma_\epsilon^2 v(\theta + \lambda). \tag{1}$$

The covariance generating functions of the two MA processes are $g_\epsilon = \sigma_\epsilon^2(1-\lambda z)(1-\lambda z^{-1})$ and $g_\delta = \sigma_\delta^2(1-z)(1-z^{-1})$, where $z = \cos(\omega) - i\sin(\omega) = e^{-i\omega}$. The covariance generating function for the aggregated process is $g_\eta = v(1+\theta z)(1+\theta z^{-1})$, where $g_\eta = g_\epsilon + g_\delta$,

$$\sigma_\epsilon^2(1-\lambda z)(1-\lambda z^{-1}) + \sigma_\delta^2(1-z)(1-z^{-1}) = v(1+\theta z)(1+\theta z^{-1}). \tag{2}$$

Let $z = 1$, and (2) becomes $\sigma_\epsilon^2(1-\lambda)^2 = v(1+\theta)^2$. Since $v^* > \sigma_\epsilon^2 + \sigma_\delta^2$, then $(1+\theta)^2 < (1-\lambda)^2$. Since $1+\theta$ and $1-\lambda$ are both positive, $\theta + \lambda < 0$. Therefore, the right hand side of (1) is positive and v^* is decreasing in λ.

The retailer follows the ConDOI policy. The order under the ConDOI policy is $(1+\Gamma\beta_0)\epsilon_t - (\lambda + \lambda\Gamma\beta_0 + \Gamma\beta_0)\epsilon_{t-1} + \lambda\Gamma\beta_0\epsilon_{t-2} + \delta_t - 2\delta_{t-1} + \delta_{t-2}$. Let $\alpha = \Gamma\beta_0/(1+\Gamma\beta_0)$. We denote the aggregate process as $S_t = \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2}$, with the covariance equations,

$$\begin{aligned} \sigma_\delta^2 + \alpha\lambda\sigma_\epsilon^2 &= \theta_2 v, \\ -4\sigma_\delta^2 - (\lambda + a)(1+\alpha\lambda)\sigma_\epsilon^2 &= \theta_1(1+\theta_2)v, \\ 6\sigma_\delta^2 + (1 + (\lambda+a)^2 + \alpha^2\lambda^2)\sigma_\epsilon^2 &= (1+\theta_1^2+\theta_2^2)v. \end{aligned} \tag{3}$$

As before, we substitute θ_1 and θ_2 with v in (3) to obtain a function $f(v,\lambda)$ in v and λ. The function $f(v,\lambda)$ satisfies$f(v,\lambda) = v^2(v+\gamma(2))^2 + v^2\gamma(1)^2 + (v+\gamma(2))^2(\gamma(2)^2 - \gamma(0)v)$, where $\gamma(2) \equiv \sigma_\delta^2 + \alpha\lambda\sigma_\epsilon^2$, $\gamma(1) \equiv -4\sigma_\delta^2 - (\lambda+a)(1+\alpha\lambda)\sigma_\epsilon^2$, $\gamma(0) \equiv 6\sigma_\delta^2 + (1+(\lambda+a)^2+\alpha^2\lambda^2)\sigma_\epsilon^2$.

We need to prove that v^* is decreasing in λ. Following the same argument from before, it is equivalent to prove $\partial f(v,\lambda)/\partial\lambda > 0$. We take derivatives of $f(v,\lambda)$ with respect to λ

$$\begin{aligned}\frac{\partial f(v,\lambda)}{\partial\lambda} &= 2(v+\gamma(2))\gamma'(2)[v^2+(\gamma(2)-\gamma(0))v+2\gamma(2)^2]+2v^2\gamma'(1)\gamma(1)-(v+\gamma(2))^2v\gamma'(0) \quad (4)\\ &= -2v^3\sigma_\epsilon^2(1+\theta_2)[\theta_1(1+2\alpha\lambda+\alpha^2)+(1+\theta_2)(\lambda+\alpha+\lambda\alpha^2)-\alpha(\theta_2+\theta_2^2-\theta_1^2)]\\ &= -2v^3\sigma_\epsilon^2(1+\theta_2)[(\alpha+\lambda-\alpha\lambda+\theta_1+\theta_2)(\alpha\theta_1+1+\theta_2)-(-\alpha\lambda+\theta_2)(1+\alpha)(\theta_1+\theta_2+1)]\end{aligned}$$

For the process $\eta_t+\theta_1\eta_{t-1}+\theta_2\eta_{t-2}$, the invertible solutions of $1+\theta_1 m+\theta_2 m^2$ lie outside the unit circle. Since $1+\theta_1 m+\theta_2 m^2=1$ at $m=0$, the function takes positive value at $m=1$, $\theta_1+\theta_2+1>0$. Since $1-\lambda-\alpha+\alpha\lambda>0$, the function $1-(\lambda+\alpha)m+\alpha\lambda m^2$ takes positive value at $m=1$.

The covariance generating functions satisfy $\sigma_\epsilon^2(1-\lambda z)(1-\alpha z)(1-\lambda z^{-1})(1-\alpha z^{-1})+\sigma_\delta^2(1-z)^2(1-z^{-1})^2=v(1+\theta_1 z+\theta_2 z^2)(1+\theta_1 z^{-1}+\theta_2 z^{-2})$. Let $z=1$, we have $\sigma_\epsilon^2(1-\lambda-\alpha+\alpha\lambda)^2=v(1+\theta_1+\theta_2)^2$. Since $v>\sigma_\epsilon^2+\sigma_\delta^2$, $(1+\theta_1+\theta_2)^2<(1-\lambda-\alpha+\alpha\lambda)^2$ and $1+\theta_1+\theta_2<1-\lambda-\alpha+\alpha\lambda$. Since $\gamma(2)>0$ and $\gamma(1)<0$, θ_2 is positive and θ_1 is negative. Since $\theta_1+\theta_2+1>0$ and $\theta_1<0$, $\alpha\theta_1+1+\theta_2>\theta_1+\theta_2+1>0$. Therefore, $(\alpha+\lambda-\alpha\lambda+\theta_1+\theta_2)(\alpha\theta_1+1+\theta_2)<0$.

We next prove $-\alpha\lambda+\theta_2>0$. If $\sigma_\delta/\sigma_\epsilon=0$, then $\theta_2=\alpha\lambda$. If $\sigma_\delta/\sigma_\epsilon\to\infty$, then $\theta_2=1$. As $\sigma_\delta/\sigma_\epsilon$ increases from 0 to ∞, θ_2 changes continuously from $\alpha\lambda$ to 1. If there exists a $\theta_2<\alpha\lambda$, there must be a $\theta_2=\alpha\lambda$ when $\sigma_\delta/\sigma_\epsilon\neq 0$. We have

$$\frac{((1+\alpha\lambda)(1-\lambda)(1-\alpha)-2(1-\alpha\lambda)^2)\alpha\lambda(1-\lambda-\alpha+\alpha\lambda)(1+\alpha\lambda)\sigma_\epsilon^2\sigma_\delta^2-(1-\alpha\lambda)^4\sigma_\delta^4}{\alpha\lambda(1+\alpha\lambda)^2(\alpha\lambda\sigma_\epsilon^2+\sigma_\delta^2)}=0.$$

Since $2>1+\alpha\lambda$, $1-\alpha\lambda>1-\lambda$ and $1-\alpha\lambda>1-\alpha$, then $(1+\alpha\lambda)(1-\lambda)(1-\alpha)<0$. Since $1-\alpha\lambda>0$, $1+\lambda-a-a\lambda>0$, $\lambda<0$ and $-(1-\alpha\lambda)^4\sigma_\delta^4<0$, the numerator is negative. Since the denominator is positive, the equation is violated. Therefore $-\alpha\lambda+\theta_2>0$.

The right hand side of (4) is positive at v^*. Following the same argument as before, v^* is decreasing in λ. ■

2 The Retailer's Demand Forecast is Optimal

In Section 3, we assume that the retailer's demand forecast L_R is a moving average of past H demands. In this section, we assume that the retailer has an optimal demand forecast. The rest assumptions are the same as in Section 3,4 and 6. We show that under this assumption, the value of information sharing is still strictly positive.

We assume that demand follows an ARMA process, $D_t=\mu+\rho_1 D_{t-1}+\rho_2 D_{t-2}+\cdots+\rho_p D_{t-p}+\epsilon_t-\lambda_1\epsilon_{t-1}-\lambda_2\epsilon_{t-2}-\cdots-\lambda_q\epsilon_{t-q}$. The optimal forecast for week $t+1$ is $\hat{D}_{t+1}=\mu+\rho_1 D_t+\rho_2 D_{t-1}+\cdots+\rho_p D_{t-p+1}-\lambda_1\epsilon_t-\lambda_2\epsilon_{t-1}-\cdots-\lambda_q\epsilon_{t-q+1}$. The optimal forecast for week $t+k$ ($k>1$) is $\hat{D}_{t,t+k}=\mu+\rho_1\hat{D}_{t+k-1}+\cdots+\rho_k D_t+\cdots+\rho_p D_{t+k-p}-\lambda_{k+1}\epsilon_{t-1}-\cdots-\lambda_q\epsilon_{t+k-q}$. Therefore, the optimal demand forecast is linear in historical demand and historical demand signals,

$$\hat{m}'_t \equiv \sum_{k=1}^{L_R} \hat{D}_{t,t+k} = \sum_{j=0}^{p-1} \beta'_j D_{t-j} + \sum_{k=0}^{q-1} c'_k \epsilon_{t-k}.$$

where β'_j is the coefficient of past j^{th} demand and c'_i is the coefficient of past i^{th}.

Under the ConDOI policy with order smoothing, the order-up-to level is $\gamma\Gamma\hat{m}'_t + (1-\gamma)I_{t-1}$. The order becomes

$$\begin{aligned} O_t &= D_t + \gamma\left(\sum_{j=0}^{p-1}\Gamma\beta'_j D_{t-j} + \sum_{k=0}^{q-1}\Gamma c'_k \epsilon_{t-k}\right) - \gamma^2\sum_{i=1}^{\infty}(1-\gamma)^{i-1}\left(\sum_{j=0}^{p-1}\Gamma\beta'_j D_{t-i-j} + \sum_{k=0}^{q-1}\Gamma c'_k \epsilon_{t-k-j}\right) \\ &\quad +\delta_t - \sum_{i=1}^{\infty}\gamma(1-\gamma)^{i-1}\delta_{t-i}. \end{aligned} \tag{5}$$

We define $\tilde{\psi}(B) \equiv 1 + \gamma\sum_{j=0}^{p-1}\Gamma\beta'_j B^j - \gamma^2\sum_{i=1}^{\infty}\sum_{j=0}^{p-1}(1-\gamma)^{i-1}\Gamma\beta'_j B^{i+j}$ as the parameter associated with demand observations, $\phi(B) \equiv \gamma\sum_{k=0}^{q-1}\Gamma c'_k B^k - \gamma^2\sum_{i=1}^{\infty}\sum_{k=0}^{q-1}(1-\gamma)^{i-1}\Gamma c'_k B^{i+k}$ as the parameter associated with demand shocks. Recall that $\kappa(B)$ is the order smoothing parameter defined in Section 4. We rewrite equation (5) as $O_t = \tilde{\psi}(B)D_t + \phi(B)\epsilon_t + \kappa(B)\delta_t$. Applying the backshift operator, we have $\pi(B)O_t = \pi(B)\tilde{\psi}(B)D_t + \pi(B)\phi(B)\epsilon_t + \kappa(B)\delta_t$, and the order becomes

$$\pi(B)O_t = \mu + \left[\varphi(B)\tilde{\psi}(B) + \pi(B)\phi(B)\right]\epsilon_t + \pi(B)\kappa(B)\delta_t,$$

where μ is the process mean, $\varphi(B)\tilde{\psi}(B) + \pi(B)\phi(B)$ is the demand shock coefficient and $\pi(B)\kappa(B)$ is the decision deviation coefficient.

Following the same spirit of Proposition 3, we need to prove that $\varphi(B)\tilde{\psi}(B) + \pi(B)\phi(B) \neq \pi(B)\kappa(B)$. Since $\tilde{\psi}(1) = 1$, $\varphi(1) \neq 0$ (due to the invertibility assumption) and $\phi(1) = 0$, we have $\varphi(1)\tilde{\psi}(1) + \pi(1)\phi(1) \neq 0$, which means the demand shock coefficient does not include the polynomial $1-B$. We already know that $\pi(B)\kappa(B)$ contains $1-B$. Henceforth, the demand shock coefficient differs from the decision deviation coefficient, and thus, the value of information sharing is strictly positive for any forecast lead time.

3 MMFE demand and GOUTP policy

In this section, we show that the value of information is strictly positive under a more general structure: the martingale model of forecast evolution (MMFE) demand and generalized order-up-to policy (GOUTP) policy studied in **?** and **?**. The MMFE model is a generalized demand model. Most time-series demand models can be interpreted as a special case of the MMFE model. The generalized order-up-to policy (GOUTP) is a stationary and affine mapping from the forecast revision to the order quantity. For the following analysis, we follow **?** and **?**'s notations.

MMFE demand. Under the MMFE structure, demand in period t is

$$D_t = \mu + \sum_{i=0}^{q} \varepsilon_{t-i,t}, \tag{6}$$

where μ is the demand mean and $\varepsilon_{t-i,t}$ is the incremental information obtained in period $t-i$ with respect to demand D_t, or more specifically, $\varepsilon_{t-i,t} = \hat{D}_{t-i,t} - \hat{D}_{t-i-1,t}$. If forecasting demand beyond q periods yields a constant prediction, then $\varepsilon_{t,t+i} = 0$ for all $i > q$. For all i, $\varepsilon_{t-i,t}$ is mutually independent, stationarily and normally distributed.

The incremental information the retailer obtains in period t, with regard to future demands, is summarized in a forecast revision vector $\boldsymbol{\varepsilon}_t = [\varepsilon_{t,t}, \varepsilon_{t,t+1}, ...\varepsilon_{t,t+q}]^T$. We assume that $\boldsymbol{\varepsilon}_t$ is independent and identically distributed with a multivariate normal $N(\mathbf{0}, \mathbf{\Sigma})$, where the variance-covariance matrix is $\mathbf{\Sigma} = E\{\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T\}$. The independently distributed $\boldsymbol{\varepsilon}_t$ implies independence between $\varepsilon_{t,t+k}$ and $\varepsilon_{t-j,t+k'}$ for any k, k' and $j > 0$.

According to the Projection Theorem, we can decompose $\varepsilon_{t,t+q}$ into $\lambda_q^0 \varepsilon_{t,t} + \varepsilon_{t,t+q}^1$ where $\varepsilon_{t,t+q}^1$ is independent of $\varepsilon_{t,t}$. We decompose $\boldsymbol{\varepsilon}_t$ into $\varepsilon_{t,t}[1, \lambda_1^0, ..., \lambda_q^0]^T + [0, \varepsilon_{t,t+1}^1, ..., \varepsilon_{t,t+q}^1]^T$. For notational convenience, let $\varepsilon_{t,t}^0$ represent $\varepsilon_{t,t}$. We then further decompose $[0, \varepsilon_{t,t+1}^1, ..., \varepsilon_{t,t+q}^1]^T$ into $\varepsilon_{t,t+1}^1[0, 1, \lambda_2^1, ..., \lambda_q^1]^T + [0, 0, \varepsilon_{t,t+2}^2, ..., \varepsilon_{t,t+q}^2]^T$. Finally, $\boldsymbol{\varepsilon}_t$ can be rewritten as

$$\boldsymbol{\varepsilon}_t \;=\; \varepsilon_{t,t}^0[1, \lambda_1^0, \lambda_2^0, ..., \lambda_q^0]^T + \varepsilon_{t,t+1}^1[0, 1, \lambda_2^1, ..., \lambda_q^1]^T + \cdots + \varepsilon_{t,t+q}^q[0, 0, ..., 0, 1]^T,$$

where $\varepsilon_{t,t+p}^p$ is independent of $\varepsilon_{t,t+q}^q$ for any $p \neq q$ and $\varepsilon_{t,t}^0$ must be nonzero. We can rewrite demand in equation (6) as $D_t = \mu + \phi^0(B)\varepsilon_{t,t}^0 + \cdots + \phi^q(B)\varepsilon_{t,t+q}^q$, where $\phi^i(B) = 1 + \lambda_1^i B + \cdots + \lambda_q^i B^q$. Demand D_t is the summation of multiple MA processes. We assume the demand is in the invertible representation. **?** shows that the aggregate MA process has an invertible representation, if at least one of the processes is invertible. Therefore, we assume the coefficient $\phi^i(B)$ is in the invertible representation for at least one $i \leq q$.

GOUTP with decision deviations. Under a GOUTP policy, the inventory level at the end of a period is a stationary and affine combination of forecast revisions

$$I_t = m + \sum_{i=0}^{q} \boldsymbol{w}_i^T \boldsymbol{\varepsilon}_{t-i}, \tag{7}$$

where m is the order mean and $\boldsymbol{w}_i$ is a weight vector defined as $\boldsymbol{w}_i = [w_{i,0}, w_{i,1}, ...w_{i,q}]^T$ for $i \geq 0$. We define $\boldsymbol{w}_{-1} = \mathbf{0}$ for notional convenience.

In practice, the order decision may depend on unobservable variables such as the transportation constraint, batching delivery and full truck load policy. As before, we introduce the error term in the order-up-to level to capture idiosyncratic shocks in decision making. We also model the decision deviation n_t as an MMFE process (see **?** for the same assumption)

$$n_t = \sum_{i=0}^{p} \delta_{t-i,t},$$

where the mean of decision deviation is zero, $\delta_{t-i,t}$ is the incremental information obtained in period $t-i$ with respect to the decision deviation n_t. p is the effective forecast horizon for decision deviations. As before, we define $\boldsymbol{\delta}_t = [\delta_{t,t}, \delta_{t,t+1}, ...\delta_{t,t+p}]^T$, which follow an $i.i.d.$ zero mean multivariate normal distribution. Following the same decomposition procedure, the decision deviation vector can be written as:

$$\boldsymbol{\delta}_t \quad = \quad \delta_{t,t}[1, \mu_1^0, \mu_2^0, ..., \mu_p^0]^T + \delta_{t,t+1}^1[0, 1, \mu_2^1, ..., \mu_p^1]^T + \cdots + \delta_{t,t+p}^p[0, 0, ..., 0, 1]^T.$$

We can rewrite decision deviations as $n_t = \varphi^0(B)\delta_{t,t}^0 + \cdots + \varphi^q(B)\delta_{t,t+q}^q$, where $\varphi^i(B) = 1 + \mu_1^i B + \cdots + \mu_q^i B^q$. The optimal mean squared forecast error implies that $\boldsymbol{\delta}_t$ and $\boldsymbol{\varepsilon}_s$ are uncorrelated for $s \neq t$ (see **?**).

The order-up-to-level is the target inventory level plus the decision deviation,

$$I_t = m + \sum_{i=0}^{q} \boldsymbol{w}_i^T \boldsymbol{\varepsilon}_{t-i} + \sum_{i=0}^{p} \delta_{t-i,t}.$$

Accordingly, the centered order takes the form:

$$O_t - \mu = \sum_{i=0}^{p+1} (\boldsymbol{e}_i - \boldsymbol{e}_{i-1})^T \boldsymbol{\delta}_{t-i} + \sum_{i=0}^{q+1} (\boldsymbol{w}_i - \boldsymbol{w}_{i-1} + \boldsymbol{e}_i)^T \boldsymbol{\varepsilon}_{t-i}, \tag{8}$$

where $\boldsymbol{e}_i$ is the unit vector with the $(i+1)th$ element equal to one.

Recall that we use B to shift variables backward in time. Note that B sets both time notations backward, $B\varepsilon_{t,t+i}^k = \varepsilon_{t-1,t+i-1}^k$. We rewrite (8) as

$$\begin{aligned} O_t - \mu \quad &= \quad [(1-B)\varphi^0(B) + \phi^0(B)]\varepsilon_{t,t}^0 + \cdots + [(1-B)\varphi^q(B) + \phi^q(B)]\varepsilon_{t,t+q}^q \qquad (9) \\ &\quad + (1-B)\left[\psi^0(B)\delta_{t,t}^0 + \psi^1(B)\delta_{t,t+1}^1 + \cdots + \psi^p(B)\delta_{t,t+p}^p\right], \\ \text{where } \varphi^i(B) \quad &= \quad \sum_{j=0}^{q} w_{0,j}\lambda_j^i + \sum_{j=0}^{q} w_{1,j}\lambda_j^i B + \cdots + \sum_{j=0}^{q} w_{q,j}\lambda_j^i B^q, \\ \phi^i(B) \quad &= \quad 1 + \lambda_1^i B + \cdots + \lambda_q^i B^q \text{ and } \psi^i(B) = 1 + \mu_1^i B + \cdots + \mu_p^i B^p. \end{aligned}$$

For notational convenience, $\lambda_j^k = 0$ for $j < k$ or $j > q$, and $\lambda_j^k = 1$ for $j = k$.

Positive value under the MMFE and GOUTP model. So far, we have decomposed orders into $p+q+2$ MA processes with respect to demand signals and decision deviations in (9),

$$\begin{aligned} X_t^i &= [\phi^i(B)+\varphi^i(B)(1-B)]\varepsilon_{t,t+i}^i, \text{ for } 0 \le i \le q, \\ Y_t^i &= \psi^i(B)(1-B)\delta_{t,t+i}^i, \text{for } 0 \le i \le p, \end{aligned} \tag{10}$$

where $\phi^i(B)+\varphi^i(B)(1-B)$ is the demand signal coefficient, and $\psi^i(B)(1-B)$ is the decision deviation coefficient. $\varepsilon_{t,t+i}^i$ is independent across t and i, $\delta_{t,t+i}^i$ is independent across t and i, and $\varepsilon_{t,t+i}^i$ is independent with $\delta_{s,s+i}^i$ for $s \neq t$. We assume that $\varepsilon_{t,t+i}^i$ is not a linear combination of $\varepsilon_{t,t+i}^{-i}$ and $\delta_{t,t+i}^j$ for any i, j, and $\delta_{t,t+i}^i$ is not a linear combination of $\delta_{t,t+i}^{-i}$ and $\varepsilon_{t,t+i}^j$ for any i, j. We normalize the coefficient of ε_t to one; thus $X_t^i = C_i^{-1}(\phi^i(B)+\varphi^i(B)(1-B))(C_i\varepsilon_t)$ where $C_i = 1+\sum_{j=0}^{L} w_{0,j}\lambda_j^i$.

The centered order becomes the aggregate process,

$$O_t - \mu = \sum_{i=0}^{q} X_t^i + \sum_{i=0}^{p} Y_t^i.$$

Since only contemporaneous demand signals and decision deviations are correlated, we can apply Theorem 2 to obtain the following proposition.

Proposition 2 *Under the MMFE demand and the GOUTP policy with decision deviations, if both demand shocks and decision deviations are nonzero, the value of information sharing is strictly positive for any finite forecast lead time $h \le \max(q,p)$.*

The sketch of the proof is as follows. According to Theorem 2, if the coefficients for any two processes are different, then the value of information sharing is positive. Since $\varphi^i(1)(1-1)=0$ for any i and $\phi^i(1) \neq 0$ for at least one i (due to the invertibility assumption), then the polynomial $1-B$ is not a factor in any demand shock coefficient. Since $\psi^j(1)(1-1)=0$ for any j, then $1-B$ is a factor in the decision shock coefficient. Therefore, the demand signals evolve differently from decision deviations, which implies that the value is strictly positive for any forecast lead time. We further strengthen our result under a more general linear and stationary demand and policy structure.

4 Empirical Results

4.1 Order Parameters

We present the estimated order parameters in Table 1. The first column records the (p,d,q) value of the ARIMA demand, and the next ten columns are the corresponding demand parameters; i.e., the 128 OR product follows an ARIMA$(3,1,0)$ demand process, $D_t - D_{t-1} = -0.72(D_{t-1}-D_{t-2}) - 0.56(D_{t-2}-D_{t-3}) - 0.31(D_{t-3}-D_{t-4}) + \epsilon_t$. For all products, demand is best estimated by $d=1$.

Table 1: Estimated order parameters.

Brand	Product	(p, d, q)	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	λ_1	λ_2	λ_3	λ_4	λ_5
Orange	128 OR	$(3, 1, 0)$	-0.72	-0.56	-0.31							
Juice			(0.10)	(0.11)	(0.10)							
	128 ORCA	$(4, 1, 0)$	-0.70	-0.63	-0.40	-0.29						
			(0.10)	(0.11)	(0.12)	(0.10)						
	12 OR	$(4, 1, 0)$	-0.88	-0.61	-0.48	-0.27						
			(0.10)	(0.13)	(0.13)	(0.10)						
	12 ORCA	$(0, 1, 1)$						0.89				
								(0.06)				
	59 ORST	$(4, 1, 0)$	-0.80	-0.57	-0.47	-0.25						
			(0.10)	(0.12)	(0.13)	(0.10)						
	59 ORPC	$(5, 1, 0)$	-0.79	-0.72	-0.57	-0.48	-0.37					
			(0.10)	(0.12)	(0.13)	(0.12)	(0.10)					
Sports	500 BR	$(2, 1, 0)$	-0.62	-0.40								
Drink			(0.09)	(0.09)								
	500 GP	$(0, 1, 5)$						0.75	0.05	0.28	0.05	-0.25
								(0.10)	(0.13)	(0.13)	(0.13)	(0.10)
	PD LL	$(0, 1, 1)$						0.57				
								(0.08)				
	PD OR	$(0, 1, 1)$						0.66				
								(0.08)				
	PD FRZ	$(0, 1, 1)$						0.66				
								(0.08)				
	1GAL GLC	$(5, 1, 0)$	-0.69	-0.57	-0.53	-0.43	-0.37					
			(0.10)	(0.11)	(0.11)	(0.11)	(0.10)					
	1GAL FRT	$(0, 1, 1)$						0.91				
								(0.05)				
	1GAL OR	$(5, 1, 0)$	-0.77	-0.52	-0.56	-0.50	-0.27					
			(0.10)	(0.12)	(0.12)	(0.12)	(0.10)					

Note. The number in parentheses denotes the standard error of the estimate.

4.2 Forecast Accuracy Summary at the Product Level

Table 2 presents the forecast accuracy percentage improvements in MAPE and MSE at the product level. We carry out two sets of comparisons: the improvement with respect to the NoInfoSharing forecast and the improvement of the InfoSharing forecast relative to the three statistical methods. The star mark means that the forecast improvement with respect to the NoInfoSharing method is statistically significant. The InfoSharing forecast in bold means a statistically significant improvement over the unbold forecasts.

Table 2 delivers two messages. First, when the supplier lacks knowledge of the replenishment policy, downstream demand information is not bringing statistically significant improvements for all products. For example, for the Reg D and O method under MAPE, only 6 products can benefit from information sharing statistically significantly. Second, incorporating the replenishment policy yields the greatest or one of the greatest improvements. The InfoSharing method has statistically higher improvement at $p < 0.1$ than all other forecast methods for 5 out 14 products under the MAPE metric and for 6 out of 14 products under the MSE metric.

Table 2: Forecast accuracy improvement summary for all methods at the product level. All methods that include downstream sales perform better than the NoInfoSharing method. The InfoSharing forecast accuracy is higher than any statistical method.

		MAPE percentage improvement				MSE percentage improvement			
Brand	Product	Vector ARIMA	Reg D	Reg D and O	Info Sharing	Vector ARIMA	Reg D	Reg D and O	Info Sharing
Orange	128 OR	11.1%	12.2%*	-14.6%	**45.0%****	8.7%	14.0%**	0.4%	**18.1%****
Juice	128 ORCA	-18.3%	8.1%	1.9%	**30.3%***	-0.5%	7.8%	**18.8%***	**26.5%****
	12 OR	31.6%	15.7%	**50.2%***	**58.6%***	32.4%**	33.8%**	35.1%**	**53.4%****
	12 ORCA	**40.8%****	**40.0%****	**38.0%****	**50.2%****	30.5%	36.3%	**57.3%****	**53.1%****
	59 ORST	**16.1%***	4.1%	5.0%	**18.8%***	13.2%*	10.9%	10.7%	7.1%
	59 ORPC	12.8%**	**29.1%****	**23.8%****	**27.7%****	**16.2%****	**31.0%****	11.4%	**29.4%****
Sports	500 BR	21.2%	26.2%	25.5%	**39.8%****	54.1%**	48.7%**	41.7%*	**62.5%****
Drink	500 GP	**30.9%***	25.7%	26.5%	**36.0%****	**53.1%****	42.9%*	38.7%*	**68.4%****
	PD LL	2.8%	-15.5%	-18.4%	4.7%	5.6%	30.9%	31.3%	51.3%
	PD OR	26.8%**	26.2%**	26.2%**	**44.2%****	43.3%*	**81.0%***	**81.0%***	**81.1%***
	PD FRZ	22.1%	8.2%	11.4%	**39.5%***	**44.5%****	8.2%	9.2%	**56.9%****
	1GAL GLC	**23.7%****	**30.3%****	**26.4%****	**38.0%****	50.1%**	42.9%*	40.1%*	**54.2%****
	1GAL FRT	**24.3%****	**21.4%***	17.2%	**29.9%****	46.4%**	40.3%**	31.3%*	**54.0%****
	1GAL OR	**16.9%***	**18.3%***	14.0%	**30.4%***	30.2%	21.2%	18.3%	**44.8%****

** 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.

Note. Significant accuracy improvement over the NoInfoSharing method is marked by a star. Significant ($p = 0.1$) accuracy improvement of the InfoSharing method over the other unbold methods is in bold.

5 Promotional Products

5.1 Theoretical Analysis

In this section, we study the impact of promotional activities on the value of information sharing. When there is a price promotion, we observe a spike in the demand during the discount activity and a slump after the activity. The growth in demand depends on the price discount rate and how long the activity lasts. Such a phenomenon might cause a non-stationary demand mean or a non-stationary covariance matrix. Therefore, we build a model to capture the promotional effect and address its impact on the value of information sharing.

According to the CPG company that we study, the supplier and the retailer pre-schedule the promotional schedule at the beginning of a year, and thus know the discount activity in advance. (There are other promotional strategies in practice. For example, **?** shows an empirical finding that inventory has significant effect on sales promotions, i.e. high inventory leads to high promotions.) Therefore, we assume an exogenous promotion schedule, where the price varies over time (a fixed price discount rate throughout a relative long period since it is equivalent to no promotion).

We assume that when there is no promotion, the underlying demand process D_t follows an ARIMA process, $\pi(B)D_t = \mu + \varphi(B)\epsilon_t$. We use X_t^P to represent the observation for promotional products and X_t to represent the baseline or the depromotionalized process, where X can be O or

D. When there is a price promotion in week t, the actual demand increases in proportion to the underlying demand,

$$D_t^P = r_t D_t, \tag{11}$$

where $r_t > 0$; we call this the promotional rate. We assume that r_t primarily depends on the week during a promotional activity (e.g. the first week of a promotional activity) and the promotional depth. We assume that the replenishment decision deviates more from the theory when the discount rate is larger (which is a reasonable assumption but is not technically required). The decision deviation becomes $r_t^\delta \delta_t$ at time t when there is a promotion, where $r_t^\delta > 0$.

As before, the retailer's demand forecast made at time t is the sum of the future L_R periods' demand forecast. The demand forecast in period k is the promotional rate in period k multiplied by the underlying demand forecast for period k,

$$\hat{m}_t = \sum_{k=1}^{L_R} r_{t+k} \hat{D}_{t,t+k}^R.$$

We choose a simple setting to illustrate the intuition where the retailer adopts the ConDOI policy with $\gamma = 1$ and $L_R = 1$. The order then becomes $r_t D_t + \Gamma r_{t+1} \hat{D}_{t,t+1}^R - \Gamma r_t \hat{D}_{t-1,t}^R$. Recall that $\hat{D}_{t,t+1}^R = \sum_{j=0}^H \beta_j D_{t-j}$ and $a_j = \Gamma \beta_j$. We rewrite the order as

$$O_t^P = r_t D_t + r_{t+1} \sum_{j=0}^H a_j D_{t-j} - r_t \sum_{j=0}^H a_j D_{t-1-j} + r_t^\delta \delta_t - r_{t-1}^\delta \delta_{t-1}.$$

We define the policy parameter in period t as $\psi_t(B) = r_t + r_{t+1} a_0 + \sum_{i=1}^{H+1} (a_i r_{t+1} - a_{i-1} r_t) B^i$, where $a_{H+1} = 0$. We define the decision deviation parameter in period t as $\kappa_t(B) = r_t^\delta - r_{t-1}^\delta B$. Both $\psi_t(B)$ and $\kappa_t(B)$ are nonstationary with the existence of promotional activities. We replace $\pi(B) D_t = \mu + \varphi(B) \epsilon_t$ in the order equation, and the order can be written as a non-stationary ARIMA process,

$$\pi(B) O_t^P = \left[r_t + (r_{t+1} - r_t) \sum_{j=0}^H a_j \right] \pi(B) \mu + \psi_t(B) \varphi(B) \epsilon_t + \pi(B) \kappa_t(B) \delta_t.$$

The order mean changes over time. We first remove the promotional lift from the order so that the process mean becomes a constant. We depromotionalize the order by dividing the promotional rate of orders $r_t^O = r_t + (r_{t+1} - r_t) \sum_{j=0}^H a_j$ in period t; then orders become

$$\pi(B) O_t \equiv \pi(B) O_t^P / r_t^O = \pi(B) \mu + \psi_t(B) \varphi(B) \epsilon_t / r_t^O + \pi(B) \kappa_t(B) \delta_t / r_t^O, \tag{12}$$

where $\psi_t(B) \varphi(B) / r_t^O$ is the coefficient associated with demand signals of degree q_ϵ and $\kappa_t(B) / r_t^O$ is the coefficient associated with decision deviations of degree q_δ.

The coefficient of ϵ_t is $(a_i r_{t+1} - a_{i-1} r_t) / (r_t + (r_{t+1} - r_t) \sum_{j=0}^H a_j)$. It changes if the promotion rate in period t differs from $t+1$. Thus, $\psi_t(B)/r_t^O$ is time-variant. If $r_t^\delta \neq r_{t+1}^\delta$, then $\kappa_t(B)$ is time-variant. The order process might not preserve the same structure as the demand in (11), since O_t^P/r_t^O might be time-variant.

When there is no information sharing, we assume the supplier can correctly estimate the promotional rate of orders r_t^O with the price and orders data. By removing the promotional lift, we obtain the baseline order process. We then use the baseline O_t to forecast the future baseline order. We measure the forecast accuracy at the baseline level (we have qualitatively equivalent results if measuring at the promotional level.)

When there is information sharing, we assume the supplier can correctly estimate the promotional rate r_t with the demand and price data and policy parameters such as a_i. Since the promotional rate depends on the week and the promotional depth, the supplier can infer the promotional rate in future periods. We then use historical demand together with estimated parameters to forecast future baseline orders.

In Section 5 of the paper, we show that the value of information sharing is strictly positive for any forecast lead time, when the retailer faces a stationary demand. The following proposition confirms the same result for nonstationary demand.

Proposition 3 *If both the demand and decision shocks are nonzero, and the retailer follows the ConDOI policy with $\gamma = 1$ and $L_R = 1$, the value of information sharing is strictly positive for any forecast finite lead time $h \leq \max\{q_\epsilon, q_\delta\}$.*

When the baseline order process is stationary, we can apply the same argument as before: the different evolution patterns of two signal series drive the positive value of information sharing. When the baseline order process is non-stationary, the baseline order yields a suboptimal predictor for the next period if the process representation changes. The value of information sharing is strictly positive even when the estimator is optimal. Therefore, we derive the same conclusion for the product with non-stationary demand.

Remark. The above result is driven by assumptions on correctly estimated parameters in both information sharing and no information sharing scenarios. The actual empirical estimation might violate this assumption. If the estimated demand and policy parameters are biased, the forecast accuracy may suffer in both cases. It then becomes unclear whether the value of sharing demand is still strictly positive. We next empirically evaluate the forecast accuracy improvement for high promotional products and display forecast results for all product lines.

5.2 Empirical Analysis

5.2.1 The empirical model without information sharing

We conduct a three-step forecasting procedure using price and order information. We first estimate the promotional lift and remove it to generate the baseline order. We then fit an ARIMA(p,d,q) model to predict the future baseline order. We finally add back the promotional lift on the predicted future baseline order.

The promotional lift is determined by the week during a promotional activity and the degree of promotional discount. Let p denote the regular price. We measure the promotional discount in time t as $\text{discount}_t = (p - p_t)/p$. We generate five week dummy variables: $Week\_Before_t$ (if t is the week before the promotional activity), $Week\_First_t$ (if t is the first week during the promotional activity), $Week\_Between_t$ (if t is one of the promotional weeks excluding the first and the last one), $Week\_Last_t$ (if t is the last promotional week) and $Week\_After_t$ (if t is the week after the promotional activity). To estimate the promotional lift, we estimate the following equation,

$$\begin{aligned}\log O_t^P &= c + a_0 Week\_Before_t \times \text{discount}_t + a_1 Week\_First_t \times \text{discount}_t \\ &\quad + a_2 Week\_Between_t \times \text{discount}_t + a_3 Week\_Last_t \times \text{discount}_t \\ &\quad + a_4 Week\_After_t \times \text{discount}_t + \varepsilon_t.\end{aligned}$$

The promotional rate of orders is $r_t^O = \exp(a_0 Week\_Before_t \times \text{discount}_t + a_1 Week\_First_t \times \text{discount}_t + a_2 Week\_Between_t \times \text{discount}_t + a_3 Week\_Last_t \times \text{discount}_t + a_4 Week\_After_t \times \text{discount}_t)$. We then generate the baseline order as $O_t = \exp(c + \varepsilon_t)$ by removing the promotional lift. We apply the NoInfoSharing forecasting method illustrated in Section 4.3 to forecast the future order $\hat{O}_{t,t+1}$. The final order forecast adds the promotional lift back $\hat{O}_{t,t+1}^P = r_{t+1}^O \hat{O}_{t,t+1}$.

5.2.2 The empirical model with information sharing

We apply the InfoSharing forecasting method illustrated in Section 4.2. We estimate the replenishment policy parameters by

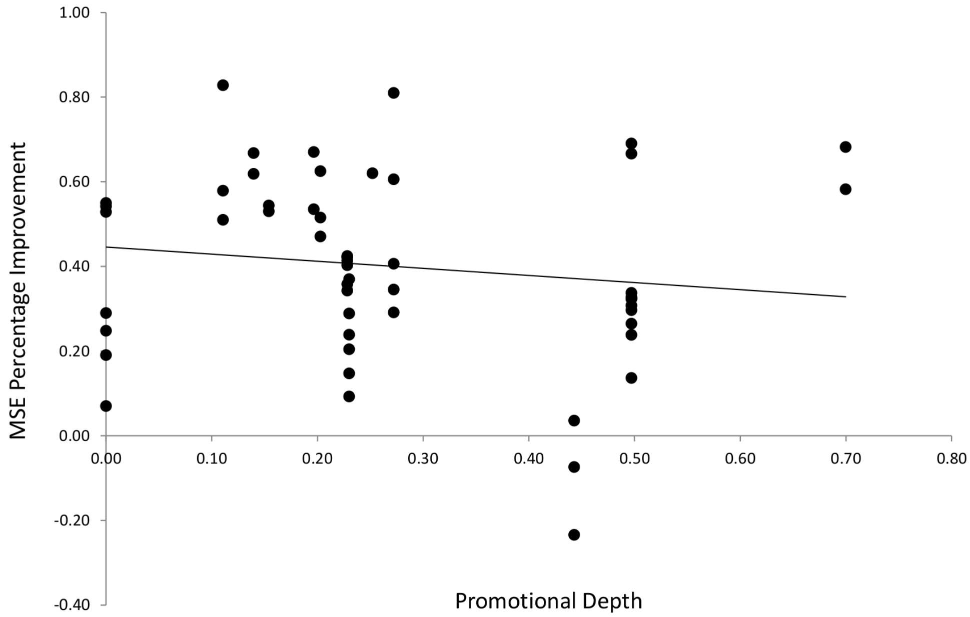
$$O_t^P = c_0 D_t^P + c_1 D_{t-1}^P + c_2 D_{t-2}^P + c_3 D_{t-3}^P - \gamma I_{t-1} + \delta_t. \qquad (13)$$

We also try to run regressions on baseline demand instead of actual demand in 13. Actual demand outperforms baseline demand in both in-sample and out-of-sample tests. To forecast the order in $t+1$, we need to forecast future demand. We follow the three-step procedure discussed above to forecast $\hat{D}_{t,t+1}^P$. The order prediction for period $t+1$ is $\hat{O}_{t,t+1}^P = c_0 \hat{D}_{t,t+1}^P + c_1 D_t^P + c_2 D_{t-1}^P + c_3 D_{t-2}^P - \gamma I_t$.

5.2.3 Empirical Results

Figure 1 displays the MSE percentage improvement with respect to the promotional depth. The points with promotional depth ≤ 0.14 correspond to our studied low-promotional products in the last column of Table 2. The first observation is that the value of information sharing is positive for most promotional products. Sharing downstream demand information is still valuable for upstream forecasts for high-promotional products. Second, we observe an insignificant correlation between forecast improvements and promotional depth. Multiple factors might impact the value differently. For example, price variations cause the order series to have a higher uncertainty, which indicates a larger room for improvement. On the other hand, the empirical model might not exactly capture the

Figure 1: Insignificant relationship between the MSE percentage improvement and the forecast lead time.



underlying dynamics of the system, and this might affect the two forecasting scenarios in different directions. If the order structure is non-stationary (in demand signals and decision deviation signals), the optimal estimators (of the ARIMA model for orders or the replenishment policy) obtained in the current week might be suboptimal for future weeks, which might affect the forecast precision of the two scenarios differently. In addition, when there is 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.

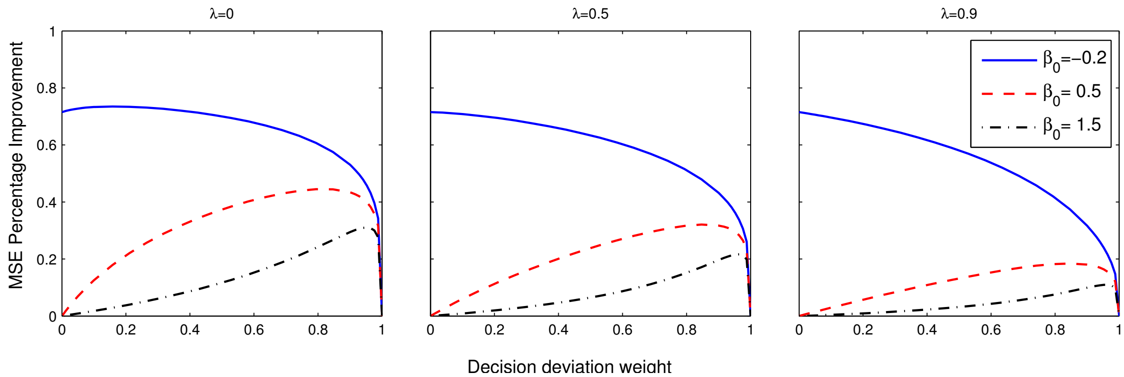
6 The Value With Respect to Policy Parameters and Forecast Lead Time

6.1 Impact of the Inventory Policy Parameters

We study the impact of the moving average weight β_0 and the DOI level Γ on the value of information sharing.

Impact of β_0. The theoretical analysis focuses on a simple setting where the retailer faces an ARIMA$(0,1,0)$ demand process and follows a ConDOI policy (it means $\gamma = 1$), since more complicated cases preclude analytically tractable solutions. Then the order process becomes $O_t = D_t + \Gamma\beta_0 D_t + \Gamma\beta_1 D_{t-1} - \Gamma\beta_0 D_{t-1} - \Gamma\beta_1 D_{t-2}$, where $\beta_0 + \beta_1 = 1$ and $\Gamma\beta_0 > -1$. Let v denote

Figure 2: Under an ARIMA(0,1,1) demand with λ and a ConDOI policy with order smoothing with $\gamma = 0.5$ and $\Gamma = 2$, the MSE percentage improvement strictly decreases in β_0



the variance of order signals. We can write the two processes associated with demand signals and decision deviations as:

$$
\begin{aligned}
X_t^1 &= (1+\Gamma\beta_0)\varepsilon_t + \Gamma(1-2\beta_0)\varepsilon_{t-1} - \Gamma(1-\beta_0)\varepsilon_{t-2}, \\
X_t^2 &= \delta_t - 2\delta_{t-1} + \delta_{t-2}.
\end{aligned}
$$

The 1-step-ahead mean squared error percentage reduction is $1 - ((1+\Gamma\beta_0)^2\sigma_\varepsilon^2 + \sigma_\delta^2)/v$. As β_0 changes, both the numerator and the denominator vary. The following proposition demonstrates that, if $\beta_0 \leq \beta_1$ or $\beta_0 \geq 0 \geq \beta_1$, the value of information sharing decreases with β_0.

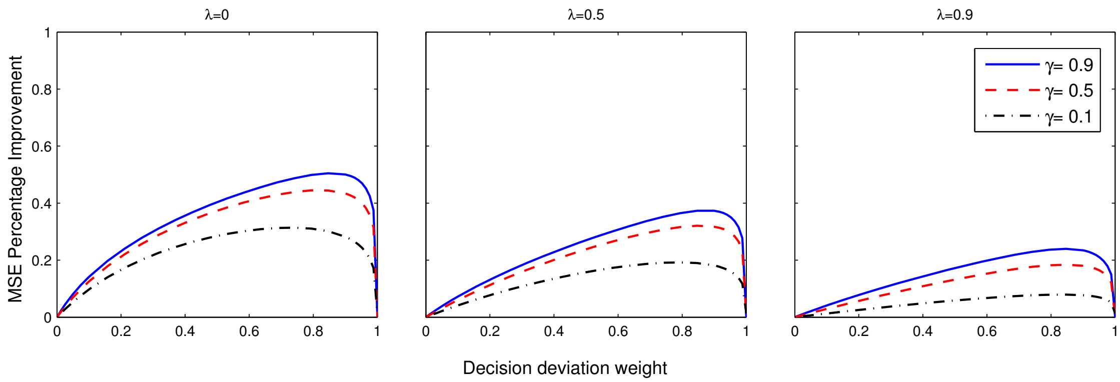
Proposition 4 *If the retailer faces an ARIMA$(0,1,0)$ demand and follows a ConDOI policy with $\beta_0 \leq \beta_1$ or $\beta_0 \geq 0 \geq \beta_1$, then the value of information sharing strictly decreases with β_0.*

Although the proposition does not cover the case of $\beta_0 > \beta_1 > 0$, we conduct a numerical study to show the same result. We present the relationship of MSE percentage improvement with respect to β_0 in Figure 2. The DOI level is 2 and the order smoothing level is 0.5. In each sub-figure, the three lines from top to bottom correspond to $\beta_0 = -0.2, 0.5$ and 1.5. The three columns from left to right correspond to $\lambda = 0, 0.5$ and 0.9.

The numerical result in Figure 2 aligns with Proposition 4. A larger weight on current week's demand means a lower benefit of including downstream demand, regardless of the decision deviation weight, demand parameters and the order smoothing level.

Impact of γ. We conduct a numerical study on γ and show its impact on the value of information sharing. We present the MSE percentage improvement with respect to γ in Figure 3. The DOI level is set to 2 and the moving average weight of demand is $(\beta_0, \beta_1) = (0.5, 0.5)$. In each sub-figure, the three lines from top to bottom correspond to $\gamma = 0.9, 0.5$ and 0.1. The three columns from left to right correspond to $\lambda = 0, 0.5$ and 0.9. Figure 3 shows that a higher order smoothing level induces a higher benefit of including downstream demand, regardless of the decision deviation weight, demand parameters and replenishment policy parameters.

Figure 3: Under an ARIMA(0,1,1) demand with λ and a ConDOI policy with order smoothing with $\beta_0 = 0.5$ and $\Gamma = 2$, the MSE percentage improvement strictly decreases in γ.



6.2 Impact of Forecast Lead Time

In this section, we numerically study the impact of forecast lead time on the forecast accuracy improvement.

We first introduce the hth-step-ahead forecast, which is defined as $\hat{S}_{t,t+h}$. The value for the hth-step-ahead forecast is positive if and only if $\mathrm{Var}(S_{t+h} - \hat{S}_{t,t+h} | \cup_i \Omega_t^{X^i}) < \mathrm{Var}(S_{t+h} - \hat{S}_{t,t+h} | \Omega_t^S)$. Recall that the h-step-ahead forecast is the sum of forecasts over the lead time, and the value is positive if and only if $\mathrm{Var}(\sum_{l=1}^h (S_{t+l} - \hat{S}_{t,t+l}) | \cup_i \Omega_t^{X^i}) < \mathrm{Var}(\sum_{l=1}^h (S_{t+l} - \hat{S}_{t,t+l}) | \Omega_t^S)$. We next study how the value of information sharing depends on lead time for these two metrics.

Figure 4 presents the MSE percentage improvement with respect to the forecast lead time for the hth-step-ahead forecast, and Figure 5 shows that for the h-step-ahead forecast. The DOI level is 2 and the order smoothing level is 0.5. The moving average weight of demand is $(\beta_0, \beta_1) = (0.5, 0.5)$. The three columns from left to right correspond to $\lambda = 0$, $\lambda = 0.5$ and $\lambda = 0.9$. Figure 4 shows that when forecasting the hth-step-ahead forecast, the value of information strictly decreases in the forecast lead time, regardless of the decision deviation weight, policy parameters and demand parameters. 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. In comparison, Figure 5 shows that when forecasting the sum of h-step-ahead forecasts, the value of information might increase in the forecast lead time under certain conditions.

7 Proofs of Propositions in the Technical Companion

Proof of Proposition 3: If either MA process $\psi_t(B)\varphi(B)\epsilon_t/r_t^O$ or $\pi(B)\kappa_t(B)\delta_t/r_t^O$ is non-stationary, and the aggregate process $\psi_t(B)\varphi(B)\epsilon_t/r_t^O + \pi^{-1}(B)\kappa_t(B)\delta_t/r_t^O$ is stationary, we can apply Theorem 2 by checking whether the coefficients of the two processes are different. We first condition the case where $\pi(1) = 0$. Then $\pi(1)\kappa_t(1)/r_t^O = 0$. Since $\psi_t(1)/r_t^O = 1$ and $\varphi(1) \neq 0$ (invertibility

Figure 4: Under an ARIMA(0,1,1) demand with λ and a ConDOI policy with order smoothing with $\gamma = 0.5, \Gamma = 2$ and $\beta_0 = 0.5$, the MSE percentage improvement of the hth-step-ahead forecast decreases in the forecast lead time.

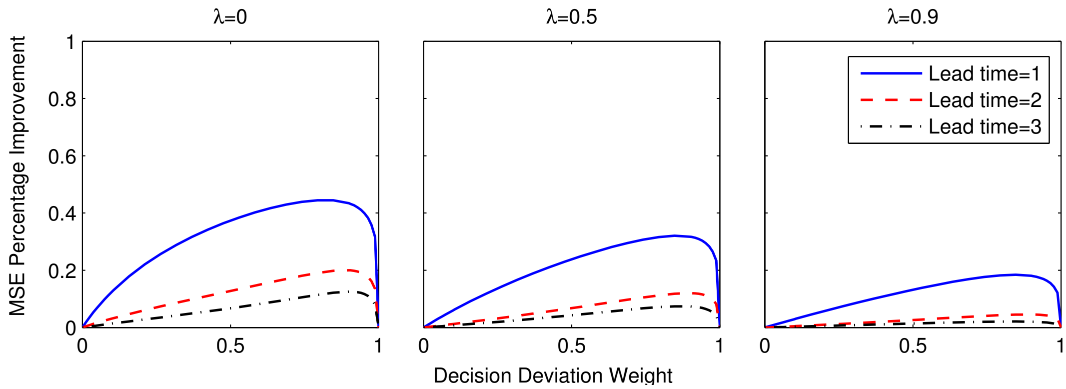
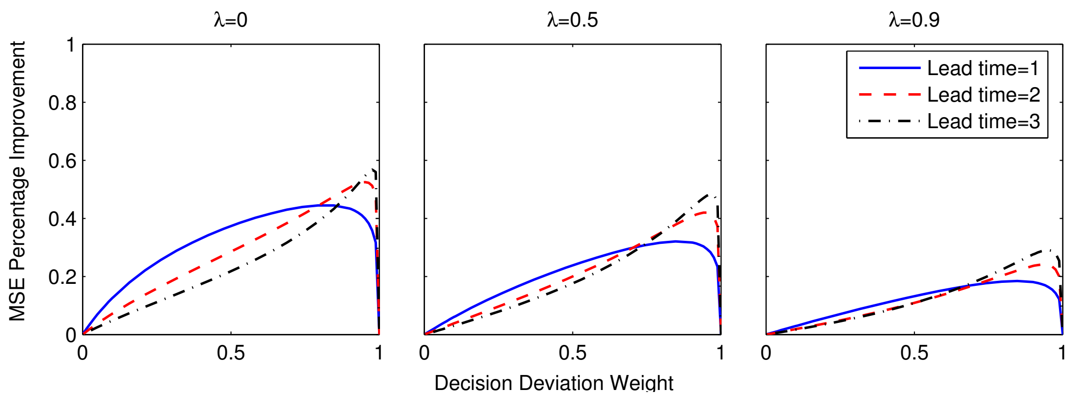


Figure 5: Under an ARIMA(0,1,1) demand with λ and a ConDOI policy with order smoothing with $\gamma = 0.5, \Gamma = 2$ and $\beta_0 = 0.5$, the MSE percentage improvement of the h-step-ahead forecast decreases in the forecast lead time.



assumption), we have $\psi_t(B)\varphi(B)/r_t^O \neq 0$. Therefore, $\pi(1)\kappa_t(1)/r_t^O \neq \psi_t(B)\varphi(B)/r_t^O$. Second, we consider another case where $\pi(B) = 1$. Then $\pi(1)\kappa_t(1)/r_t^O = (r_t^\delta - r_{t-1}^\delta)/r_t^O$. We know that $\psi_t(1)\varphi(1)/r_t^O = \varphi(1)$. Since for promotional products, the sign of $r_t^\delta - r_{t-1}^\delta$ at the beginning of a promotional activity differs from when an activity ends, then $\pi(1)\kappa_t(1)/r_t^O \neq \psi_t(1)\varphi(1)/r_t^O$ for some period t. According to Theorem 2, the value of information sharing is strictly positive.

We next analyze the case where any of the two processes is non-stationary and the aggregate process $\psi_t(B)\varphi(B)\epsilon_t/r_t^O + \pi(B)\kappa_t(B)\delta_t/r_t^O$ is non-stationary. With information sharing, since the supplier knows the promotional rate r_{t+1} at time t, we can apply the detailed order structure, and thus the forecast error has the least unresolved uncertainty. Without information sharing, the ARIMA estimator of orders obtained in the current period might be suboptimal for future periods. The future order forecast error might be enlarged by such non-optimality.■

Proof of Proposition 4: Recall that the aggregate process (or the order process) is $S_t =$

$\eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2}$. We denote $(1+a_0)^2\sigma_\epsilon^2 + \sigma_\delta^2$ as v_{Share}. The MSE percentage improvement is $(v - v_{Share})/v$. Recall that $a_1 = \beta_1\Gamma$ and $a_0 = \beta_0\Gamma$. We prove v_{Share}/v is increasing in a_0 or equivalently v/v_{Share} is decreasing in a_0. Let v' denote v/v_{Share}.

The aggregate MA process satisfies the covariance equations

$$\begin{aligned}
\left[\sigma_\delta^2 - (1+a_0)(\Gamma - a_0)\sigma_\epsilon^2\right]/v_{Share} &= \theta_2 v' \\
\left[-4\sigma_\delta^2 + (1+2a_0-\Gamma)(\Gamma-2a_0)\sigma_\epsilon^2\right]/v_{Share} &= \theta_1(1+\theta_2)v' \\
\left[6\sigma_\delta^2 + ((1+a_0)^2 + (\Gamma-2a_0)^2 + (\Gamma-a_0)^2)\sigma_\epsilon^2\right]/v_{Share} &= (1+\theta_1^2+\theta_2^2)v'
\end{aligned} \tag{14}$$

Substituting θ_1 and θ_2 with v', we derive a function $f(v', a_0)$ with variable v' and parameter a_0,

$$f(v, a_0) = v^2\left(v + \gamma(2)\right)^2 + v^2\gamma(1)^2 + \left(v + \gamma(2)\right)^2\left(\gamma(2)^2 - \gamma(0)v\right),$$

where $\gamma(2) = [\sigma_\delta^2 - (1+a_0)(\Gamma-a_0)\sigma_\epsilon^2]/v_{Share}$, $\gamma(1) = [-4\sigma_\delta^2 + (1+2a_0-\Gamma)(\Gamma-2a_0)\sigma_\epsilon^2]/v_{Share}$ and $\gamma(3) = [6\sigma_\delta^2 + ((1+a_0)^2 + (\Gamma-2a_0)^2 + (\Gamma-a_0)^2)\sigma_\epsilon^2]/v_{Share}$. The goal is to prove the invertible parameter v' of $f(v', a_0) = 0$ is decreasing in a_0, which is equivalent to prove $\partial f(v', a_0)/\partial a_0 > 0$.

We first take derivatives of $f(v', a_0)$ with respect to a_0,

$$\frac{\partial f(v', a_0)}{\partial a_0} = 2(v' + \gamma(2))\gamma'(2)\left[v'^2 + (\gamma(2) - \gamma(0))v' + 2\gamma(2)^2\right] + 2v'^2\gamma'(1)\gamma(1) - (v' + \gamma(2))^2 v'\gamma'(0).$$

We then substitute $\gamma'(2)$, $\gamma'(1)$ and $\gamma'(0)$ with θ_1 and θ_2 from the covariance equations,

$$\begin{aligned}
\frac{\partial f(v', a_0)}{A\partial a_0} &= \theta_1\left[(4\Gamma - 8a_0 - 2)v_{Share} - 2(a_0+1)(-4\sigma_\delta^2 + (1+2a_0-\Gamma)(\Gamma-2a_0)\sigma_\epsilon^2)\right] \\
&\quad + (1+\theta_2)\left[(3\Gamma - 6a_0 - 1)v_{Share} + 2(a_0+1)(6\sigma_\delta^2 + ((1+a_0)^2 + (\Gamma-2a_0)^2 + (\Gamma-a_0)^2)\sigma_\epsilon^2)\right] \\
&\quad + (\theta_2 + \theta_2^2 - \theta_1^2)\left[(2a_0 + 1 - \Gamma)v_{Share} - 2(a_0+1)(\sigma_\delta^2 - (1+a_0)(\Gamma-a_0)\sigma_\epsilon^2)\right] \\
&= \sigma_\delta^2\left[(\theta_1+\theta_2+1)(\theta_2-\theta_1-3)(-a_0-a_1-1) + 2(\theta_1+\theta_2+1) + 6(1+\theta_2)(1+a_0)\right] \\
&\quad + \sigma_\epsilon^2\left[(a_0+1)^2(\theta_1+\theta_2+1)(\Gamma+1)(\theta_2-\theta_1+1) + B\right]
\end{aligned}$$

where $A = 2v'^3\sigma_\epsilon^2(1+\theta_2)v_{Share}^{-2}$ and $B = 2(a_0+1)(1+\theta_2)(-(a_0-a_1)(a_1+1)+a_1^2) - 2(a_0+1)\theta_1(a_0-a_1)(a_0+a_1+2) - 2(a_0+1)\theta_1(1+a_1)$. The coefficient of v_{Share} is obviously positive. The derivation is positive if and only if the coefficient of σ_ϵ^2 and σ_δ^2 are positive.

Positive σ_δ^2. Since $a_0 + a_1 = \Gamma$, then $-a_0 - a_1 - 1$ is negative. Since $\theta_2 - \theta_1 - 3$ is also negative, $6(1+\theta_2)(1+a_0)$ and $2(\theta_1+\theta_2+1)$ are positive, the coefficient of σ_δ^2 is positive.

Positive σ_ϵ^2. $(a_0+1)^2(\theta_1+\theta_2+1)(a_0+a_1+1)(\theta_2-\theta_1+1)$ is positive. The coefficient of $1+\theta_2$ is larger than that of θ_1, $-(a_0-a_1)(a_1+1)+a_1^2 > -(a_0-a_1)(a_0+a_1+2) - (1+a_1)$.

If $a_0 \le a_1$, then $-(a_0-a_1)(a_1+1)+a_1^2$ is positive. Recall that $1+\theta_2 > \theta_1$. If θ_1 and $-(a_0-a_1)(a_0+a_1+2)-(1+a_1)$ are both positive or both negative, then B is positive. If $\theta_1 < 0$ and $-(a_0-a_1)(a_0+a_1+2)-(1+a_1) > 0$, then $B \ge 2(a_0+1)(\theta_1+\theta_2+1)((a_1-a_0)^2 + a_1^2 + (1+a_0-a_1)(1+a_1)) > 0$. If $\theta_1 > 0$ and $-(a_0-a_1)(a_0+a_1+2)-(1+a_1) < 0$, then since $1+\theta_2 > \theta_1$,

$B > 0$.

If $a_1 < -1$, then $-(a_0 - a_1)(a_1 + 1) + a_1^2$ is positive. Since $-(a_0 - a_1)(a_0 + a_1 + 2) - (1 + a_1) < -2a_0 + a_1 - 1$, which is negative, $-(a_0 - a_1)(a_0 + a_1 + 2) - (1 + a_1)$ is negative. Since θ_1 is negative and $1 + \theta_2$ is positive, $B > 0$.

If $-1 < a_1 \leq 0 \leq a_0$, we rearrange the coefficient of σ_ϵ^2 to $(a_0 + 1)(\theta_1 + \theta_2 + 1)[(\Gamma - a_1 + 1)(\Gamma + 1)(\theta_2 - \theta_1 + 1) + 2a_1^2 - 2(\Gamma - 2a_1)(a_1 + 1)] + 2\theta_1(a_0 + 1)[-(a_0 - a_1)(a_0 + 1) - 1 - a_1 - a_1^2]$. Since $-(a_0 - a_1)(a_0 + 1) - 1 - a_1 - a_1^2 < 0$ and $\theta_1 < 0$, the second part is positive. Since $a_1 < 0$, $\theta_2 > 0$ and $\theta_1 < 0$. We then have $\theta_2 - \theta_1 + 1 > 1$. As a result, $(\Gamma - a_1 + 1)(\Gamma + 1)(\theta_2 - \theta_1 + 1) + 2a_1^2 - 2(\Gamma - 2a_1)(a_1 + 1) \geq 6a_1^2 - 3(\Gamma - 1)a_1 + 1 + \Gamma^2$. Since $-6\Gamma - 5(1 + \Gamma^2) < 0$, then $6a_1^2 - 3(\Gamma - 1)a_1 + 1 + \Gamma^2$ is positive. Therefore, the coefficient of σ_ϵ^2 is positive.■