Aggregated Information in Supply Chains

Vladimir Kovtun*, Avi Giloni†, Clifford Hurvich ‡,

February 8, 2018

Abstract

We study a two-stage supply chain where the retailer observes two demand streams coming from two consumer populations. We further assume that each demand sequence is a stationary Autoregressive Moving Average (ARMA) process with respect to a Gaussian white noise sequence (shocks). The shock sequences for the two populations could be contemporaneously correlated. We show that it is typically optimal for the retailer to construct its order to its supplier based on forecasts for each demand stream (as opposed to the sum of the streams) and that doing so is never sub-optimal. We demonstrate that the retailer's order to its supplier is ARMA and yet can be constructed as the sum of two ARMA order processes based upon the two populations. When there is no information sharing, the supplier only observes the retailer's order which is the aggregate of the two aforementioned processes. In this paper, we determine when there is value to sharing the retailer's individual orders, and when there is additional value to sharing the retailer's individual shock sequences. We also determine the supplier's mean squared forecast error under no sharing, process sharing, and shock sharing.

*Sy Syms School of Business, Yeshiva University, Suite 334, 245 Lexington Avenue, New York, NY 10012. E-mail: *vladimir.kovtun@yu.edu*

†Sy Syms School of Business, Yeshiva University, BH-428, 500 West 185th St., New York, NY 10033. E-mail: *agiloni@yu.edu*

‡Department of Information, Operations, and Management Science, Leonard N. Stern School of Business, New York University, Suite 8-52, 44 West 4th St., New York, NY 10012. E-mail: *churvich@stern.nyu.edu*

KEYWORDS: Supply Chain Management, Information Sharing, Time Series, ARMA, Invertibility, Demand Sharing, Process Sharing, Order-up-to policy.

1 Introduction

In this paper we consider a two-stage supply chain where the retailer observes two demand streams coming from two consumer populations. We further assume that each demand sequence is a stationary Autoregressive Moving Average (ARMA) process with respect to a Gaussian white noise sequence (shocks). The shock sequences for the two populations can be contemporaneously correlated. We show that in the vast majority of cases it is optimal for the retailer to construct its order to its supplier based on forecasts for each demand stream (as opposed to the sum of the streams) and that doing so is never sub-optimal. When there is no information sharing between the retailer and the supplier, the supplier observes the aggregated order placed by the retailer. The contribution of our paper is that we determine when there is value to the retailer sharing the two individual order processes (process sharing) and when there is additional value to the retailer sharing its two individual shock sequences (shock sharing). We also determine the supplier's mean squared forecast error (MSFE) under no sharing, process sharing, and shock sharing.

There has been much research on the value of information sharing in a supply chain. Lee et al [7] (hereafter referred to as LST), Raghunathan [8], Zhang [11], Gaur et al. [3] (hereafter referred to as GGS), Giloni et al [4] (hereafter referred to as GHS), and Kovtun et al [6] (hereafter referred to as KGH) studied the value of information sharing in supply chains under AR and ARMA demand. Zhang and GGS extend the original work of LST and Raghunathan by studying the value of information sharing in supply chains where the retailer serves an ARMA(p,q) demand as opposed to AR(1) demand. In each of these papers, the retailer places orders with a supplier using a periodic review order-up-to policy. Both the supplier and the retailer know the parameters of the demand process; however, the retailer may or may not choose to share information about the actual realizations of demand with the supplier. Zhang studied how the order process propagates upstream in a supply chain under the assumption that ARMA demand to the retailer and all upstream players is invertible. In such a case, there would be no value of information sharing to any of the players.

GGS were first to point out that the retailer's order to the supplier may not be invertible (i.e., the current shock cannot be obtained as a linear combination of present and past demand observations only) even though the retailer's demand is invertible. In other words, the supplier's demand may not be invertible with respect to the retailer's shocks, even though the retailer's demand is invertible with respect to its own shocks. GHS characterized the supplier's best linear forecast with and without information sharing from the retailer. They showed that the retailer's order to the supplier is QUARMA (quasi-ARMA) and characterized the value of information sharing when each supply chain player determines its best linear forecast of lead time demand using all of its available information (depending upon sharing arrangements). KGH studied the propagation of demand in a supply chain where a supply chain player may share its demand, its shocks, or nothing at all with the immediate upstream player. They compared the three sharing scenarios and characterized when demand sharing is superior to no sharing and when shock sharing is superior to demand sharing.

Overall, there are two key elements to accurately determining the value of information in the aforementioned research. First, it is essential to determine when the retailer's order to its supplier is invertible with respect to its shocks and when it is not. Second, when the retailer's order is not invertible with respect to its shocks, it is essential to determine the supplier's best linear forecast in this case. GHS appears to have been the first to do so and we utilize and extend the framework considered by GHS and KGH in this paper.

Our specific research problem is most closely related to a recent paper by Cui et al [2] (hereafter referred to as Cui) who studied a supply chain where a supplier receives orders that are an aggregate of two processes due to the retailer placing an order that is generated by a standard inventory policy coupled with order smoothing and a decision deviation process. In other words, in their paper, the existence of the two processes is due to the manner in which the planner operates. They concluded that sharing the demand processes is almost always valuable. Our results match those of Cui when the retailer's order or actually retailer's orders are invertible with respect to its shock sequence(s). However, when the retailer's orders are not invertible with respect to its shock sequences, there exist additional cases where process sharing is not valuable. We further consider when the retailer may share its shocks with the supplier. Finally, we demonstrate how to compute the one-step ahead mean squared forecast error (MSFE). In other words, besides describing whether or not there is value to information sharing, we also quantify this value as determined by the difference in the MSFE under sharing compared to the no sharing arrangement. We further demonstrate these differences among the three situations that we study: no sharing, process sharing, and shock sharing.

The mathematical problem of determining whether there is value of process sharing (in terms of reducing the MSFE) was considered and solved by Kohn [5] under the key assumptions that the bivariate system is in its *Wold representation* and that the univariate aggregated process is also in its Wold representation. A Wold representation expresses the series as a linear combination of present and past shocks, where the shocks contain precisely all of the information retrievable from the linear past of the (univariate or bivariate) series[1]. We note that a series represented as a linear combination of present and past shocks is in its Wold representation if and only if it is invertible with respect to its shocks. Following this methodology, we consider the two demand

[1] We are implicitly assuming that the series is purely nondeterministic as would be the case for any ARMA model

sequences at the retailer and its order to the supplier as a sum of two ARMA sequences which are bivariate ARMA. In a simpler context without any aggregation, GHS demonstrated that the retailer's order can naturally become non-invertible with respect to the retailer's shocks even though the retailer's demand is invertible with respect to its shocks. Thus, one cannot assume that the retailer's aggregated order to the supplier nor the bivariate ARMA system representing the retailer's order processes is invertible with respect to its shocks and hence Kohn's key assumptions may not apply to our problem.

2 Problem Setup

In this section, we describe the mathematical problem at hand. To begin, we assume that the retailer observes ARMA demand from two populations and hence its demand sequences $\{D_{1,t}\}$ and $\{D_{2,t}\}$ are ARMA with respect to two (possibly contemporaneously correlated) shock sequences $\{\tilde{\epsilon}_{1,t}\}$ and $\{\tilde{\epsilon}_{2,t}\}$. Using the backshift operator B (where $BD_t = D_{t-1}$), the retailer's two demand streams are

$$\Phi_1(B)D_{1,t} \quad = \quad d_1 + \Theta_1^\star(B)\tilde{\epsilon}_{1,t} \tag{1}$$

$$\Phi_2(B)D_{2,t} \quad = \quad d_2 + \Theta_2^\star(B)\tilde{\epsilon}_{2,t} \tag{2}$$

where $\Theta_1^\star(B) = 1 - \theta_{1,1}^\star B - \ldots - \theta_{1,q_1}^\star B^{q_1}$, $\Theta_2^\star(B) = 1 - \theta_{2,1}^\star B - \ldots - \theta_{2,q_2}^\star B^{q_2}$, and $\Phi_1(B) = 1 - \phi_{1,1}B - \ldots - \phi_{1,p_1}B^{p_1}$, $\Phi_2(B) = 1 - \phi_{2,1}B - \ldots - \phi_{2,p_2}B^{p_2}$. We assume that $\Phi_1(z)$ and $\Phi_2(z)$ have all their roots outside the unit circle and that $\Theta_1^\star(z)$ and $\Theta_1^\star(z)$ have all their roots either outside or on the unit circle. These conditions insure that $\{D_{1,t}\}$ and $\{D_{2,t}\}$ are stationary and invertible with respect to $\{\tilde{\epsilon}_{1,t}\}$ and $\{\tilde{\epsilon}_{2,t}\}$ respectively. Invertibility insures that $\{\tilde{\epsilon}_{1,t}\}$ and $\{\tilde{\epsilon}_{2,t}\}$ are observable to the retailer at time t, meaning that the retailer can obtain $\tilde{\epsilon}_{1,t}$ and $\tilde{\epsilon}_{2,t}$ using elements only in $\{D_{1,n}\}_{n=-\infty}^{t}$ and $\{D_{2,t}\}_{n=-\infty}^{t}$ respectively. This model is reasonable under the assumption that $\{D_{1,t}\}$ does not Granger cause $\{D_{2,t}\}$ and $\{D_{2,t}\}$ does not Granger cause $\{D_{1,t}\}$. Furthermore this implies that $\{(\tilde{\epsilon}_{1,t}, \tilde{\epsilon}_{2,t})'\}$ are the shocks appearing in the Wold representation of the bivariate series $\{(D_{1,t}, D_{2,t})'\}$ which could be expressed using the bivariate ARMA model

$$\begin{pmatrix} \Phi_1(B) & 0 \\ 0 & \Phi_2(B) \end{pmatrix} \begin{pmatrix} D_{1,t} \\ D_{2,t} \end{pmatrix} = \begin{pmatrix} \Theta_1^\star(B) & 0 \\ 0 & \Theta_2^\star(B) \end{pmatrix} \begin{pmatrix} \tilde{\epsilon}_{1,t} \\ \tilde{\epsilon}_{2,t} \end{pmatrix}, \tag{3}$$

or equivalently using the Wold representation

$$\begin{pmatrix} D_{1,t} \\ D_{2,t} \end{pmatrix} = \begin{pmatrix} \Psi_1^\star(B) & 0 \\ 0 & \Psi_2^\star(B) \end{pmatrix} \begin{pmatrix} \tilde{\epsilon}_{1,t} \\ \tilde{\epsilon}_{2,t} \end{pmatrix} \tag{4}$$

where $\Psi_1^\star(z) = \dfrac{\Theta_1^\star(z)}{\Phi_1(z)}$ and $\Psi_2^\star(z) = \dfrac{\Theta_2^\star(z)}{\Phi_2(z)}$.

We note that the retailer's total demand at time $t+1$ is given by $D_{1,t+1} + D_{2,t+1}$. The retailer could forecast this demand using $\{D_{1,n} + D_{2,n}\}_{n=-\infty}^{t}$ or it can add the two forecasts of $D_{1,t+1}$ and $D_{2,t+1}$ based upon the two individual sequences $\{D_{1,n}\}_{n=-\infty}^{t}$ and $\{D_{2,n}\}_{n=-\infty}^{t}$ respectively. The question of whether it is better to use the individual processes for forecasting has been answered in Theorem 1 of Kohn [5].

Remark 1 *The best linear forecast based on* $\{D_{1,n}\}_{n=-\infty}^{t}$ *and* $\{D_{2,n}\}_{n=-\infty}^{t}$ *is better than the best linear forecast based on* $\{D_{1,n}+D_{2,n}\}_{n=-\infty}^{t}$ *if and only if* $\dfrac{\Theta_1^\star(z)}{\Phi_1(z)} \neq \dfrac{\Theta_2^\star(z)}{\Phi_2(z)}$ *for some* $z \in \mathbb{C}$*. Otherwise the best linear forecasts based on* $\{D_{1,n}\}_{n=-\infty}^{t}$ *and* $\{D_{2,n}\}_{n=-\infty}^{t}$ *or on* $\{D_{1,n} + D_{2,n}\}_{n=-\infty}^{t}$ *are equivalent.*

This remark follows directly from the aforementioned theorem, which states (in slightly more generality) that given a Wold representation of a bivariate system $\{(x_t, y_t)'\}$,

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \sum_{j=0}^{\infty} A(j) \begin{pmatrix} e_{1,t-j} \\ e_{2,t-j} \end{pmatrix}, \tag{5}$$

where the $A(j)$ are 2×2 matrices, $A(0) = I$, the 2×2 identity matrix, and $\{(e_{1,t}, e_{2,t})'\}$ are the Wold shocks which generate the same linear past as $\{(x_t, y_t)'\}$, the two best linear forecasts of $x_{t+1} + y_{t+1}$ obtained using $\{x_n + y_n\}_{n=-\infty}^{t}$ or using $\{x_n\}$ and $\{y_n\}$ will be equivalent if and only if there exist scalar constants k_j such that

$$A(j)' \begin{pmatrix} 1 \\ 1 \end{pmatrix} = k_j \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ for all } j \geq 1. \tag{6}$$

Otherwise, the forecast based on $\{x_n\}$ and $\{y_n\}$ will be optimal. Applying this to representation (2) we obtain Remark 1.

In order to study the value of information sharing to the supplier, we first need to obtain an expression for the retailer's order, which we assume is determined by a myopic order-up-to-policy. The retailer's demand at time $t+1$ is given by

$$D_{t+1} = D_{1,t+1} + D_{2,t+1}. \tag{7}$$

In light of Remark 1, we consider $m_{1,t}$ and $m_{2,t}$, which are the best linear forecasts of leadtime demand for each demand stream, based on information for the given stream available at time t . If indeed, the retailer had observed demand from the first demand stream, then its order would be $X_t = D_{1,t} + m_{1,t} - m_{1,t-1}$ and similarly if the retailer had observed demand from only the second demand stream, its order would be $Y_t = D_{2,t} + m_{2,t} - m_{2,t-1}$. Hence $m_t = m_{1,t} + m_{2,t}$ is the best linear forecast of the retailer's leadtime demand. It can be easily shown that $Var\Big((D_{1,t+1} + D_{2,t+1})|\mathcal{M}_t\Big) = Var(\epsilon_{1,t}) + Var(\epsilon_{2,t}) + 2Cov(\epsilon_{1,t}, \epsilon_{2,t})$ where $\mathcal{M}_t$ is all the information available to the retailer at time t. It follows that the retailer's myopic order-up-to-level determined using information available at time t is given by

$$\begin{aligned} S_t &= m_t + c\sqrt{Var(\epsilon_{1,t}) + Var(\epsilon_{2,t}) + 2Cov(\epsilon_{1,t}, \epsilon_{2,t})} \\ &= m_{1,t} + m_{2,t} + c\sqrt{Var(\epsilon_{1,t}) + Var(\epsilon_{2,t}) + 2Cov(\epsilon_{1,t}, \epsilon_{2,t})}. \end{aligned}$$

where c is the retailer's required service level given by $c = \Phi^{-1}[\frac{p}{p+h}]$ where h and p are holding and shortage costs and Φ is the standard Normal cdf. The resulting order is then

$$Z_t = D_t + S_t - S_{t-1} \tag{8}$$

or simply,

$$Z_t = D_t + m_t - m_{t-1} = D_{1,t} + m_{1,t} - m_{1,t-1} + D_{2,t} + m_{2,t} - m_{2,t-1} = X_t + Y_t. \tag{9}$$

We refer to $\{X_t\}$ and $\{Y_t\}$ as the retailer's order processes. From Theorem 1 of GHS, it follows that each of $\{X_t\}$ and $\{Y_t\}$ are quasi-ARMA with respect to each of the retailer's shock sequences with the same AR polynomials $\Phi_1(z)$ and $\Phi_2(z)$ appearing in (1) and (2). For simplicity in this paper, we assume that $\{X_t\}$ and $\{Y_t\}$ are ARMA with respect to these shocks. Hence we are considering the case that the retailer's order to the supplier is given by $Z_t = X_t + Y_t$, where

$$\Phi_1(B)X_t = d_1 + \Theta_1(B)\epsilon_{1,t}, \tag{10}$$

$$\Phi_2(B)Y_t = d_2 + \Theta_2(B)\epsilon_{2,t}, \tag{11}$$

and $\Theta_1(B) = 1 - \theta_{1,1}B - \ldots - \theta_{1,q_1}B^{q_1}$ and $\Theta_2(B) = 1 - \theta_{2,1}B - \ldots - \theta_{2,q_2}B^{q_2}$ are the resulting MA polynomials based upon the propagation described above with $\epsilon_{1,t} = \lambda_1\tilde{\epsilon}_{1,t}$ and $\epsilon_{2,t} = \lambda_2\tilde{\epsilon}_{2,t}$ for constants λ_1 and λ_2. A constructive algorithm for obtaining polynomials $\Theta_1(z)$, $\Theta_2(z)$ and constants λ_1, λ_2 from polynomials $\Phi_1(z)$, $\Phi_2(z)$, $\Theta_1^\star(z)$ and $\Theta_2^\star(z)$ is provided in Theorem 3 of KGH.

In matrix notation, we represent $\{(X_t, Y_t)'\}$ as a bivariate ARMA process

$$\begin{pmatrix} \Phi_1(B) & 0 \\ 0 & \Phi_2(B) \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \Theta_1(B) & 0 \\ 0 & \Theta_2(B) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \tag{12}$$

with $\Sigma_\epsilon = Cov[(\epsilon_{1,t}, \epsilon_{2,t})'] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$, and $\sigma_1^2\sigma_2^2 - \sigma_{12}^2 > 0$, where we have assumed that $d_1 = d_2 = 0$ for notational simplicity.

The bivariate process $\{(X_t, Y_t)'\}$ is causal with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ if $det \begin{pmatrix} \Phi_1(z) & 0 \\ 0 & \Phi_2(z) \end{pmatrix} \neq 0$ for all z such that $|z| \leq 1$. Similarly, $\{(X_t, Y_t)'\}$ is invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ if $det \begin{pmatrix} \Theta_1(z) & 0 \\ 0 & \Theta_2(z) \end{pmatrix} \neq 0$ for all z such that $|z| \leq 1$. Since each of the processes is assumed to be causal with respect to each individual shock series, it follows that the bivariate process is also causal with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$.

GHS however note that it is possible for an MA polynomial (such as $\Theta_1(z)$) to have a root inside the unit circle. Hence $\{X_t\}$ will not be invertible with respect to $\{\epsilon_{1,t}\}$. In this case the bivariate process will not be invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ as well. When this is the case, it means that at time t it is not possible to obtain $(\epsilon_{1,t}, \epsilon_{2,t})'$ from the linear past of $\{(X_t, Y_t)'\}$.

Nonetheless, in all situations we will consider in this paper, it is possible to represent $\{(X_t, Y_t)'\}$ with respect to an observable shock sequence $\{(e_{1,t}, e_{2,t})'\}$, where $(e_{1,t}, e_{2,t})'$ is the difference between $(X_t, Y_t)'$ and the best linear forecast of $(X_t, Y_t)'$ at time $t-1$ using the infinite past of $\{(X_t, Y_t)'\}$. This shock sequence appears in the unique Wold representation of $\{(X_t, Y_t)'\}$, which exists for any stationary time series. We thus refer to $\{(e_{1,t}, e_{2,t})'\}$ as the Wold shocks.

We note that $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ in equation (12) will not in general be the Wold shocks, specifically when $\{(X_t, Y_t)'\}$ is not invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$. GHS explained that the supplier cannot forecast its demand using shocks that are not observable to the supplier even though they are observable to the retailer unless the retailer shares its shocks with the supplier. The same applies here except that like in KGH we contrast the cases of process sharing, shock sharing, and no sharing.

When $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ is not observable to the supplier we consider the Wold representation of the retailer's bivariate order system given by

$$\left(\begin{array}{c} X_t \\ \\ Y_t \end{array}\right) = \left(\begin{array}{cc} \Psi_{11}(B) & \Psi_{12}(B) \\ \\ \Psi_{21}(B) & \Psi_{22}(B) \end{array}\right)\left(\begin{array}{c} e_{1,t} \\ \\ e_{2,t} \end{array}\right), \tag{13}$$

Wold shocks which generate the same linear past as $\{(X_t, Y_t)'\}$ and $\Psi_{11}(z) = \sum_{j=0}^{\infty} \psi_{11j} z^j$, $\Psi_{12}(z) = \sum_{j=0}^{\infty} \psi_{12j} z^j$, $\Psi_{21}(z) = \sum_{j=0}^{\infty} \psi_{21j} z^j$, and $\Psi_{22}(z) = \sum_{j=0}^{\infty} \psi_{22j} z^j$ are (potentially) infinite degree polynomials, with $\Psi_{12}(z) = \Psi_{21}(z) = 0$ when the system is diagonal. Unless specified otherwise, for polynomials $P(z)$ and $Q(z)$, the equivalence $P(z) = Q(z)$ should be interpreted to be for all complex-valued z in this paper.

where $\{(e_{1,t}, e_{2,t})'\}$ are the We note that although the bivariate ARMA in (12) has a diagonal MA matrix[2], the Wold representation of the bivariate system need not be diagonal (see Theorem 3). In other words, even though we assume the bivariate system that describes the retailer's order to the supplier has a diagonal MA matrix, a nondiagonal Wold representation of this system can organicadeqully occur. An immediate consequence of this is that it would not be possible to apply the methodology of GHS and KGH to evaluate information sharing, since univariate representations cannot accurately describe a non-diagonal bivariate representation.

Similarly, consider the Wold representation of the retailer's order to the supplier, which is actually the supplier's demand (see GHS who discuss the Order-Demand Non-Equivalence Property):

$$Z_t = \gamma_t + \sum_{j=1}^{\infty} a_j \gamma_{t-j}, \tag{14}$$

where $\{\gamma_t\}$ are the univariate Wold shocks, i.e., those shocks that span the linear past of $\{Z_t\}$ and are thus observable by the supplier when there is no information sharing. If there is no additional information shared by the retailer, the supplier uses (14) to forecast its leadtime demand.

[2]Throughout this paper, when expressing the process $\{(X_t, Y_t)'\}$ with respect to a set of shocks, we refer to the matrix of MA coefficients on the right-hand side of the expression as the MA matrix.

We study when there is value to the supplier for the retailer to share the values of its processes, X_t and Y_t. In other words, we study whether this sharing arrangement would lead to a superior forecast of its demand than the one provided by (14). We also study when there is additional value to the retailer sharing the values of its shocks, $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$. Finally, we describe how to compute the one step ahead MSFE under the three arrangements, (i) no sharing, (ii) process sharing, and (iii) shock sharing.

Kohn's formulation discussed previously can be used to study when the MSFE of the best linear forecast of Z_{t+1} given the history of $\{Z_t\}$ is equivalent to the MSFE of the best linear forecast of Z_{t+1} given the history of $\{(X_t, Y_t)'\}$. It is important to note that his results are based upon the knowledge of the polynomials and coefficients therein with respect to the Wold shocks. Thus, in order to apply Kohn's results, one needs to first obtain the Wold representation of the bivariate series $\{(X_t, Y_t)'\}$. For example, from Theorem 1 of Kohn, it follows that there is no value to sharing $(X_t, Y_t)'$ if and only if there exist scalar constants k_j such that

$$\begin{pmatrix} \psi_{11j} & \psi_{12j} \\ \psi_{21j} & \psi_{22j} \end{pmatrix}' \begin{pmatrix} 1 \\ 1 \end{pmatrix} = k_j \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ for all } j \geq 1. \tag{15}$$

It follows from the same theorem that the constants k_j in (15) will be the constants a_j in (14). We note that according to Condition (15), if $A(j)$ is diagonal for all j, then there is no value to sharing $\{(X_t, Y_t)'\}$ if and only if $\psi_{11j} = \psi_{22j}$ for all j. Using Kohn's approach where possible, the roadmap of the remainder of the paper is as follows.

- In Section 3.1, (Invertible Case) we consider the value of information sharing when the retailer's two processes, $\{X_t\}$ and $\{Y_t\}$ are invertible ARMA processes with respect to the retailer's shocks. We provide a necessary and sufficient condition under which there is no value to process sharing (or shock sharing).

- In Section 3.2, (Noninvertible Case with mutually independent shocks) we consider when at least one of $\{X_t\}$ and/or $\{Y_t\}$, each ARMA, is not invertible with respect to the retailer's shocks and these two shock sequences are mutually independent. We provide a necessary and sufficient condition under which there is no value to process sharing.

- In Section 3.3, (Noninvertible Case with contemporaneously correlated shocks) we consider when $\{(X_t, Y_t)'\}$ is a bivariate MA(q) such that the series is not invertible with respect to the retailer's shocks and these shock are contemporaneously correlated. We show the Wold representation of a bivariate ARMA (and therefore MA) system is no longer diagonal. Using the algorithm developed by Tunicliffe-Wilson [10] we find the Wold representation and obtain several additional examples of no value to process sharing under these conditions.

- In Section 4 (MSFE) we obtain the one-step-ahead forecasts and their corresponding MSFEs under the three sharing arrangements for the scenarios in Section 3.1 - Section 3.3.

3 The Value of Information Sharing

In this section we study the value of process sharing and shock sharing when $\{(X_t, Y_t)'\}$ can be modeled as a bivariate ARMA(p, q) given by (12), which is a natural consequence of the two ARMA demand streams observed by the retailer. Evaluating process sharing and comparing its value to that of shock sharing is dependent on whether $\{(X_t, Y_t)'\}$ is invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ as well as on whether Σ_ϵ in (12) is diagonal. Therefore, we proceed by considering these different cases in the subsections below. Details on how to compute the one-step ahead MSFE for the cases discussed here can be found in Section 4.

3.1 Invertible Case

Here we assume that $\{(X_t, Y_t)'\}$ is invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, meaning that *det* $\begin{pmatrix} \Theta_1(z) & 0 \\ 0 & \Theta_2(z) \end{pmatrix} \neq 0$ for all z such that $|z| \leq 1$ in (12). Since the MA matrix is assumed to be diagonal, this condition is equivalent to stating that $\Theta_1(z)$ and $\Theta_2(z)$ have no roots inside the unit circle. This means that we require the individual univariate processes $\{X_t\}$ and $\{Y_t\}$ to be invertible with respect to $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$ respectively and do not need to directly consider the bivariate system. It follows that $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ are the Wold shocks of $\{(X_t, Y_t)'\}$. Furthermore, forecast errors, as well as corresponding MSFEs are identical under shock sharing and process sharing (see Remark 2 below).

Remark 2 *If $\{(X_t, Y_t)'\}$ is invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ then forecasts of elements in $\{Z_n\}_{n=t+1}^{\infty}$ are identical under process sharing and shock sharing since all elements in $\{(X_n, Y_n)'\}_{n=-\infty}^{t}$ can be recovered from $\{(\epsilon_{1,n}, \epsilon_{2,n})'\}_{n=-\infty}^{t}$ and vice-versa.*

From the remark above we note that the main question posed in this subsection is whether process sharing (and shock sharing) are valuable when compared with no sharing. This is addressed in the following theorem which establishes when forecasts of Z_{t+1} under no sharing will be equivalent to forecasts under either of the two sharing arrangements.

Theorem 1 *Suppose the retailer observes processes which can be modeled using (12) with det* $\begin{pmatrix} \Theta_1(z) & 0 \\ 0 & \Theta_2(z) \end{pmatrix} \neq 0$ *for all z such that $|z| \leq 1$. There is no value to sharing $\{X_t\}$ and $\{Y_t\}$ (or $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$) if and only if* $\dfrac{\Theta_1(z)}{\Phi_1(z)} = \dfrac{\Theta_2(z)}{\Phi_2(z)}$.

The proof of Theorem 1, which can be found in the Appendix, is based upon Theorem 1 of Kohn [5]. Note that Theorem 1 of this paper, which holds under the invertibility of $\{(X_t, Y_t)'\}$ with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, matches the result described by Cui for this case when considering decision deviations. In the following subsections we discuss why this equivalence does not hold in general and only follows from this invertibility assumption. If there are no common roots between $\Theta_1(z)$ and $\Phi_1(z)$ and no common roots between $\Theta_2(z)$ and $\Phi_2(z)$ then the condition that $\frac{\Theta_1(z)}{\Phi_1(z)} = \frac{\Theta_2(z)}{\Phi_2(z)}$ amounts to checking whether $\Theta_1(z) = \Theta_2(z)$ and $\Phi_1(z) = \Phi_2(z)$.

Although it may appear that the likelihood of there being no value to information sharing is small, this is not the case as highlighted by the following remark.

Remark 3 *If the retailer's two demand streams $\{D_{1,t}\}$ and $\{D_{2,t}\}$ are MA(1) in (1) and (2) and the retailer's leadtime is 0 (orders arrive in the next period), then there is no value to sharing the individual order processes (or the individual shock sequences).*

The assumptions of the remark guarantee that the processes $\{X_t\}$ and $\{Y_t\}$ will be white noise and therefore the conditions of Theorem 1 will hold. To see this, suppose $\{D_t\}$ is MA(1) with respect to a white noise sequence $\{\tilde{\epsilon}_t\}$ such that

$$D_t = \tilde{\epsilon}_t - \theta_1^{\star}\tilde{\epsilon}_t \tag{16}$$

From Theorem 3 of KGH, this induces orders $\{X_t\}$ such that

$$X_t = (1 - \theta_1^{\star})\tilde{\epsilon}_t \tag{17}$$

which we could rewrite with respect to white noise sequence $\{\epsilon_t\}$ as

$$X_t = \epsilon_t \tag{18}$$

where $\epsilon_t = (1 - \theta_1^{\star})\tilde{\epsilon}_t$.

We provide several examples in the Appendix which illustrate how demand propagates from the retailer and supplier and how information sharing becomes valuable for the invertible case described in Theorem 1. In the following subsections we study the non-invertible case.

3.2 Non-Invertible Case with Mutually Independent Shocks

In this subsection, we study whether information sharing is valuable in the event that the retailer's process $\{(X_t, Y_t)'\}$ is noninvertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, meaning that in (12) there exists $|z_0| < 1$ such that $det \begin{pmatrix} \Theta_1(z_0) & 0 \\ 0 & \Theta_2(z_0) \end{pmatrix} = 0$. Since the MA matrix is diagonal, this is equivalent to requiring $\Theta_1(z)$ or $\Theta_2(z)$ to have at least one root inside the unit circle. As indicated by the previous subsection, one of the determinants of the value of process and shock sharing arrangements is the ARMA representation of $\{(X_t, Y_t)'\}$ with respect to the Wold shocks of $\{(X_t, Y_t)'\}$. By the assumptions considered here, $\{(X_t, Y_t)'\}$ is not invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ and therefore $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ in (12) are not the Wold shocks of $\{(X_t, Y_t)'\}$. As we will show below, the form of the Wold shocks depends on whether the shock series $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$ are mutually independent or contemporaneously correlated. The former assumption is the basis for this subsection.

Proposition 1 provides the ARMA representation of $\{(X_t, Y_t)'\}$ with respect to their Wold shocks under the assumption that the shock series are independent, i.e., that $\sigma_{12} = 0$ in Σ_ϵ. The proposition is given below following a necessary definition.

Definition 1 *Suppose that $\Theta(z)$ is a polynomial of order q in a complex variable z, with leading coefficient 1 and roots $\{r_k\}$. Suppose also that none of the roots r_k are on the unit circle. Then*

$$\Theta(z) = \Pi_{k=1}^{q}(1 - z/r_k).$$

Let IN denote the set of roots such that $|r_k| < 1$. Let OUT denote the set of roots such that $|r_k| > 1$. Define

$$\Theta^\dagger(z) = \prod_{r_k \in IN}(1 - zr_k) \prod_{r_k \in OUT}(1 - z/r_k).$$

where IN is the list of all roots, r_k, of $\Theta(z)$ with repeated entries to allow for multiplicities, such that $|r_k| < 1$.

We note that $\Theta^{\dagger}(z)$ has all of its roots outside the unit circle. If $\Theta(z)$ has no roots inside the unit circle, then $\Theta(z) = \Theta^{\dagger}(z)$.

Proposition 1 *Suppose the retailer's processes can be modeled by (12), where there exists* $|z_0| < 1$ *such that* $\det \begin{pmatrix} \Theta_1(z_0) & 0 \\ 0 & \Theta_2(z_0) \end{pmatrix} = 0$ *and* $\sigma_{12} = 0$. *An ARMA representation of the retailer's processes with respect to its Wold shocks* $\{(e_{1,t}, e_{2,t})'\}$ *is given by*

$$\begin{pmatrix} \Phi_1(B) & 0 \\ 0 & \Phi_2(B) \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \Theta_1^{\dagger}(B) & 0 \\ 0 & \Theta_2^{\dagger}(B) \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}. \tag{19}$$

where $\Sigma_e = cov[(e_{1,t}, e_{2,t})'] = \begin{pmatrix} \sigma_1^2 \prod_{r_j \in IN1} |r_j|^{-2} & 0 \\ 0 & \sigma_2^2 \prod_{r_j \in IN2} |r_j|^{-2} \end{pmatrix}$ *where* $IN1$ *(and* $IN2$*) is the list of all roots,* r_j*, of* $\Theta_1(z)$ *(and* $\Theta_2(z)$*) with repeated entries to allow for multiplicities, such that* $|r_j| < 1$.

The proof of Proposition 1, along with an instructive lemma can be found in the Appendix. We note that the mutual independence of the shocks in (12) guarantees that the bivariate ARMA representation of $\{(X_t, Y_t)'\}$ with respect to the Wold shocks $\{(e_{1,t}, e_{2,t})'\}$ has a diagonal MA and AR matrix. Furthermore, determining $\{(e_{1,t}, e_{2,t})'\}$ and the Wold representation of $\{(X_t, Y_t)'\}$ is equivalent to determining the Wold shocks $\{e_{1,t}\}$ and $\{e_{2,t}\}$ for each univariate series $\{X_t\}$ and $\{Y_t\}$.

Due to the noninvertibility of $\{(X_t, Y_t)'\}$ with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, shock sharing and process sharing result in different forecasts and different MSFEs. Extending the results by GHS and KGH for univariate processes, if either $\Theta_1(z)$ or $\Theta_2(z)$ have at least one root inside the unit circle in (12) then the MSFE when forecasting Z_{t+1} at time t is always smaller under shock sharing than under process sharing. The difference in forecasts stems from the fact that it is possible to recover elements in $\{(X_n, Y_n)'\}_{n=-\infty}^{t}$ using $\{(\epsilon_{1,n}, \epsilon_{2,n})'\}_{n=-\infty}^{t}$ but not vice-versa. Hence $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ contains more information than $\{(X_t, Y_t)'\}$.

Since (19) has the same form as (12) and $\{(X_t, Y_t)'\}$ is invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ we can apply Theorem 1 to this representation to assess the value of process sharing. When the AR and MA matrices in the ARMA representation of $\{(X_t, Y_t)'\}$ with respect to the Wold shocks are diagonal, as in Proposition 1, the value of process sharing (over no sharing) rests on whether $\frac{\Theta_1^\dagger(z)}{\Phi_1(z)}$ and $\frac{\Theta_2^\dagger(z)}{\Phi_2(z)}$ are equivalent as discussed in Theorem 2 below.

Theorem 2 *Suppose the retailer observes processes modeled by (12), where the covariance matrix of* $(\epsilon_{1,t}, \epsilon_{2,t})'$ *is given by* $\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. *There is no value to process sharing if and only if* $\frac{\Theta_1^\dagger(z)}{\Phi_1(z)} = \frac{\Theta_2^\dagger(z)}{\Phi_2(z)}$.

The proof follows immediately by applying Theorem 1 to model (19) given by Proposition 1. According to Theorem 2, there is no value to process sharing if $\Theta_1(z) = \Theta_2(z)$ and $\Phi_1(z) = \Phi_2(z)$. This can be seen by noting that if $\Theta_1(z) = \Theta_2(z)$ then $\Theta_1^\dagger(z) = \Theta_2^\dagger(z)$. However this is not the only possibility of there being no value to process sharing as the condition of Theorem 2 can still be met while $\Theta_1(z) \neq \Theta_2(z)$ but $\Phi_1(z) = \Phi_2(z)$.[3] The phenomenon observed here can be explained as follows. Even though the original ARMA representations of X_t and Y_t are such that $\frac{\Theta_1(z)}{\Phi_1(z)} \neq \frac{\Theta_2(z)}{\Phi_2(z)}$, the ratios of the AR and MA polynomials in the ARMA representations corresponding to the Wold shocks could be equal.

For example suppose $\Phi_1(z) = \Phi_2(z) = 1$ and $\{Y_t\}$ is invertible with respect to $\{\epsilon_{2,t}\}$ and hence $\Theta_2^\dagger(B) = \Theta_2(B)$. On the other hand, if at least one root of $\Theta_1(z)$ lies inside the unit circle and yet $\Theta_1^\dagger(z) = \Theta_2(z) = \Theta_2^\dagger(z)$, it follows that there is no value to process sharing. This is demonstrated

[3]In other words, this demonstrates that Theorem 2 of Cui is incomplete in the sense that there are additional cases under which there is no value to process sharing.

in the example below when $\{(X_t, Y_t)'\}$ is bivariate MA(1).

Example 1 *Suppose $\{D_{1,t}\}$ and $\{D_{2,t}\}$ are each MA(2) with respect to $\{\tilde{\epsilon}_{1,t}\}$ and $\{\tilde{\epsilon}_{2,t}\}$ respectively, given by*

$$D_{1,t} = (1 - .2B - .4B^2)\tilde{\epsilon}_{1,t} \tag{20}$$

$$D_{2,t} = (1 - 1.2B + .4B^2)\tilde{\epsilon}_{2,t}. \tag{21}$$

We note that $D_{1,t}$ and $\{D_{2,t}\}$ are invertible with respect to $\{\tilde{\epsilon}_{1,t}\}$ and $\{\tilde{\epsilon}_{2,t}\}$ respectively. From Theorem 3 of KGH, $\{(X_t, Y_t)'\}$ can be represented as the bivariate ARMA process

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - .5B & 0 \\ 0 & 1 - 2B \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{22}$$

where $\epsilon_{1,t} = (1 - .2)\tilde{\epsilon}_{1,t}$ and $\epsilon_{2,t} = (1 - 1.2)\tilde{\epsilon}_{2,t}$. Without loss of generality we assume that $\sum_\epsilon = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

We note that $\{(X_t, Y_t)'\}$ is causal but not invertible with respect to $\{\epsilon_{1,t}, \epsilon_{2,t}\}$ since $det \begin{pmatrix} 1 - .5z & 0 \\ 0 & 1 - 2z \end{pmatrix} = 0$ for $z = .5$. By Definition 1, since $\Theta_2(z) = 1 - 2z$, it follows that $\Theta_2^\dagger(z) = 1 - .5z$. From Proposition 1, the bivariate ARMA representation of $\{(X_t, Y_t)'\}$, which is invertible with respect to $\{e_{1,t}, e_{2,t}\}$, is given by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - .5B & 0 \\ 0 & 1 - .5B \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}. \tag{23}$$

By applying Theorem 2, there is no value to process sharing in this case since $\Phi_1 = \Phi_2 = 1$ and $\Theta_1^\dagger = \Theta_2^\dagger = 1$, even though it may have seemed like there would be value based on the model in (22).

It is also possible to find many examples of no value to information sharing when $\Phi_1 \neq \Phi_2$. For instance, let $\Phi_1(z) = \Phi_2(z)(1 - .5z)$ and $\Theta_1(z) = \Theta_2(z)(1 - 2z)$ where $\Theta_2(z)$ has all its roots outside the unit circle. We can note that $\Theta_1^\dagger(z) = \Theta_2(z)(1-.5z)$, $\Theta_2^\dagger = \Theta_2$ and indeed $\dfrac{\Theta_1^\dagger}{\Phi_1} = \dfrac{\Theta_2^\dagger}{\Phi_2}$. These examples are structured such that $\Theta_1^\dagger(z)$ cancels root(s) of $\Phi_1(z)$. As a demonstration, consider the following.

Example 2 *Let $\{(X_t, Y_t)'\}$ be bivariate ARMA(2,1):*

$$\begin{pmatrix} 1-.7B+.01B^2 & 0 \\ 0 & 1-.2B \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1-2B & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \tag{24}$$

where $\sum_\epsilon$ is given by $\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$.

Although at first glance it may appear that there would be value to information sharing in this case since $\Phi_1 \neq \Phi_2$ and $\Theta_1 \neq \Theta_2$, we note that $\{(X_t, Y_t)'\}$ is not invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ and thus proceed to determine the ARMA representation of $\{(X_t, Y_t)'\}$ with respect to the Wold shocks. From Prop 1, the bivariate ARMA representation of $\{(X_t, Y_t)'\}$, which is invertible with respect to $\{e_{1,t}, e_{2,t}\}$, is given by

$$\begin{pmatrix} 1-.7B+.01B^2 & 0 \\ 0 & 1-.2B \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1-.5B & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}. \tag{25}$$

Noting that $\dfrac{\Theta_1^\dagger(z)}{\Phi_1(z)} = \dfrac{1-.5z}{1-.7z+.01z^2} = \dfrac{1-.5z}{(1-.2z)(1-.5z)} = \dfrac{1}{1-.2z} = \dfrac{\Theta_2^\dagger(z)}{\Phi_2(z)}$ we use Theorem 2 to determine that there is no value to process sharing.

We now explore the impact of information sharing on the supplier's MSFE when $\{(X_t, Y_t)'\}$ is not invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$. In the discussion below, we make use of Theorem 4, Corollary 2 and Theorem 5 of Section 4 to compute MSFEs under shock sharing, process sharing and no sharing respectively.

Example 3 *Suppose* $\{D_{1,t}\}$ *and* $\{D_{2,t}\}$ *are each MA(2) with respect to* $\{\tilde{\epsilon}_{1,t}\}$ *and* $\{\tilde{\epsilon}_{2,t}\}$ *respectively, given by*

$$D_{1,t} = (1 - .2B - .4B^2)\tilde{\epsilon}_{1,t} \tag{26}$$

$$D_{2,t} = (1 - 1.2B - \theta^{\star}_{2,2}B^2)\tilde{\epsilon}_{2,t}. \tag{27}$$

where $-1 < \theta^{\star}_{2,2} < -.2$. *We note that this restriction guarantees that* $\{D_{2,t}\}$ *is invertible with respect to* $\{\tilde{\epsilon}_{2,t}\}$. *It follows from Theorem 3 of KGH, that* $\{(X_t, Y_t)'\}$ *can be represented as the bivariate ARMA process*

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - .5B & 0 \\ 0 & 1 - \dfrac{\theta^{\star}_{2,2}}{1-1.2}B \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{28}$$

where $\epsilon_{1,t} = (1 - .2)\tilde{\epsilon}_{1,t}$ *and* $\epsilon_{2,t} = (1 - 1.2)\tilde{\epsilon}_{2,t}$. *Without loss of generality we assume that* $\sum_{\epsilon} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. *The determinant of* $\begin{pmatrix} 1 - .5z & 0 \\ 0 & 1 - \dfrac{\theta^{\star}_{2,2}}{1-1.2}z \end{pmatrix}$ *has a root inside the unit circle when* $|\theta^{\star}_{2,2}| > .2$.

In Figure 1, the ratio of MSFE under no sharing to the MSFE under process sharing is shown when $-1 < \theta^{\star}_{2,2} < -.2$. Note that the MSFE under no sharing is equal to the MSFE under process sharing if and only if $\theta^{\star}_{2,2} = -.4$, such that $\theta_{2,2} = 2$ and $\Theta^{\dagger}_2(z) = 1 - .5z$ as is expected due to Theorem 2.

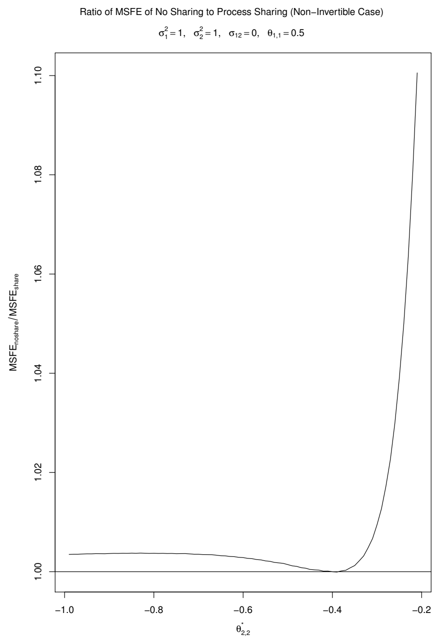

Figure 1: The one-step ahead MSFE ratio of no sharing to process sharing is shown for the parameters in (28). We note in this case $\theta_{1,1} = .5$ and that there is no value to process sharing when $\theta^{\star}_{2,1} = -.4$, such that $\theta_{2,2} = 2$.

Comparing this with Figure 2, where the ratio of MSFE under no sharing and MSFE under shock sharing is shown, one observes that shock sharing is always valuable. In general, the scales of Figure 1 and 2 indicate that process sharing has much less value than shock sharing. We note that the larger the modulus of $\theta^{\star}_2$, the farther the root of $\Theta_2(z)$ inside the unit circle, making shock sharing increasingly more valuable.

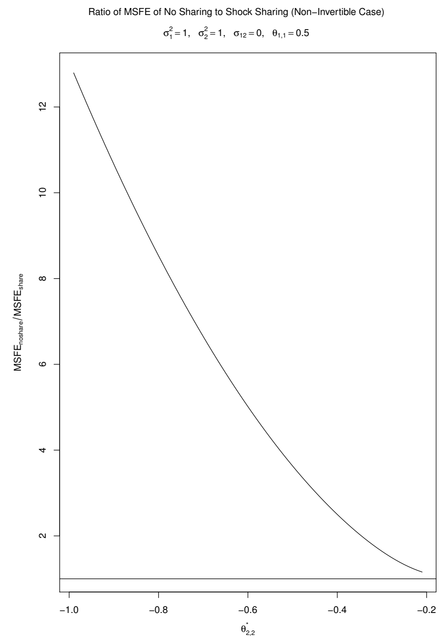

Figure 2: The one-step ahead MSFE ratio of no sharing to shock sharing is shown for the parameters in (28). A horizontal line is drawn at 1, corresponding to no value to shock sharing.

3.3 Bivariate Non-Invertible Diagonal MA(q) Process with Contemporaneously Correlated Shocks

In this subsection we consider the case where $\{(X_t, Y_t)'\}$ is a noninvertible bivariate MA(q) with respect to shock sequences, $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, that are contemporaneously correlated. That is, we now consider the retailer's processes where

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \Theta_1(B) & 0 \\ 0 & \Theta_2(B) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \tag{29}$$

with $\Theta_1(B) = 1 - \theta_{1,1}B - \ldots - \theta_{1,q_1}B^{q_1}$, $\Theta_2(B) = 1 - \theta_{2,1}B - \ldots - \theta_{2,q_2}B^{q_2}$ such that $max(q_1, q_2) = q$ and there exists $|z_0| < 1$ such that $det \begin{pmatrix} \Theta_1(z_0) & 0 \\ 0 & \Theta_2(z_0) \end{pmatrix} = 0$. Here we assume that the shock covariance matrix Σ_ϵ (defined under (12)) is non-diagonal and thus $\sigma_{12} \neq 0$. The last assumption implies that it is no longer sufficient to simply invert each univariate process in order to obtain the Wold representation.

In Theorem 3 below we show that a bivariate ARMA(p,q) process (see (12)) with (i) diagonal AR and MA matrices, (ii) a non-diagonal covariance matrix of $(\epsilon_{1,t}, \epsilon_{2,t})'$, (iii) where $\{(X_t, Y_t)'\}$ is non-invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, has a bivariate ARMA$(p,q)$ representation with respect to the shocks appearing in its Wold representation where the MA matrix is strictly non-diagonal. We note that in this section we are interested in the value of information sharing only for bivariate MA(q) processes. Nonetheless Theorem 3 below applies to bivariate MA(q) processes as well as the more general bivariate ARMA(p,q) case.

Theorem 3 *Suppose the retailer's processes are given by (12) and let $IN1$ and $IN2$ be the list of roots of $\Theta_1(z)$ and $\Theta_2(z)$ inside the unit circle respectively such that $|IN1| + |IN2| \geq 1$. If $IN1 \neq IN2$ and the covariance matrix of the shocks is nondiagonal ($\sigma_{12} \neq 0$), then the MA matrix in the Wold representation of $\{(X_t, Y_t)'\}$ is nondiagonal.*

Proof. Consider the covariance matrix $\Gamma(h) = E((X_{t+h}, Y_{t+h})'(X_t, Y_t))$ for $h \in \mathbb{Z}$. The covariance matrix generating function, defined as $G(z) = \sum_{h=-\infty}^{\infty} \Gamma(h) z^h$, of system (12) is given by (see Brockwell and Davis [1] pp 420, Equation (11.3.17))

$$\begin{pmatrix} \Phi_1^{-1}(z) & 0 \\ 0 & \Phi_2^{-1}(z) \end{pmatrix} \begin{pmatrix} \Theta_1(z) & 0 \\ 0 & \Theta_2(z) \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \Theta_1(\frac{1}{z}) & 0 \\ 0 & \Theta_2(\frac{1}{z}) \end{pmatrix} \begin{pmatrix} \Phi_1^{-1}(\frac{1}{z}) & 0 \\ 0 & \Phi_2^{-1}(\frac{1}{z}) \end{pmatrix} \tag{30}$$

$$= \begin{pmatrix} \Phi_1^{-1}(z)\Theta_1(z)\sigma_1^2\Theta_1(\frac{1}{z})\Phi_1^{-1}(\frac{1}{z}) & \Phi_1^{-1}(z)\Theta_1(z)\sigma_{12}\Theta_2(\frac{1}{z})\Phi_2^{-1}(\frac{1}{z}) \\ \Phi_2^{-1}(z)\Theta_2(z)\sigma_{12}\Theta_1(\frac{1}{z})\Phi_1^{-1}(\frac{1}{z}) & \Phi_2^{-1}(z)\Theta_2(z)\sigma_2^2\Theta_2(\frac{1}{z})\Phi_2^{-1}(\frac{1}{z}) \end{pmatrix}. \tag{31}$$

We will now use the fact that the covariance matrix generating function corresponding to the Wold representation must be equivalent to (31) to present a proof by contradiction. Suppose the Wold representation of $\{(X_t, Y_t)'\}$ is given by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} A_1(B) & 0 \\ 0 & A_2(B) \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}, \tag{32}$$

where $A_1(z) = \sum_{k=0}^{\infty} a_{1,k} z^k$ and $A_2(z) = \sum_{k=0}^{\infty} a_{2,k} z^k$ and the covariance of the Wold shocks $\begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}$ is given by $\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$. The covariance matrix generating function of this system is given by

$$\begin{pmatrix} A_1(z) & 0 \\ 0 & A_2(z) \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} A_1(\frac{1}{z}) & 0 \\ 0 & A_2(\frac{1}{z}) \end{pmatrix} \tag{33}$$

$$= \begin{pmatrix} A_1(z)\alpha A_1(\frac{1}{z}) & A_1(z)\beta A_2(\frac{1}{z}) \\ A_2(z)\beta A_1(\frac{1}{z}) & A_2(z)\gamma A_2(\frac{1}{z}) \end{pmatrix} = \begin{pmatrix} \Phi_1^{-1}(z)\Theta_1(z)\sigma_1^2\Theta_1(\frac{1}{z})\Phi_1^{-1}(\frac{1}{z}) & \Phi_1^{-1}(z)\Theta_1(z)\sigma_{12}\Theta_2(\frac{1}{z})\Phi_2^{-1}(\frac{1}{z}) \\ \Phi_2^{-1}(z)\Theta_2(z)\sigma_{12}\Theta_1(\frac{1}{z})\Phi_1^{-1}(\frac{1}{z}) & \Phi_2^{-1}(z)\Theta_2(z)\sigma_2^2\Theta_2(\frac{1}{z})\Phi_2^{-1}(\frac{1}{z}) \end{pmatrix}. \tag{34}$$

The polynomials $\Theta_1(z)$ and $\Theta_2(z)$ can be expressed as

$$\Theta_1(z) = \prod_{r_i \in OUT1} (1 - r_i^{-1} z) \prod_{r_j \in IN1} (1 - r_j^{-1} z), \tag{35}$$

$$\Theta_2(z) = \prod_{r_i \in OUT2} (1 - r_i^{-1} z) \prod_{r_j \in IN2} (1 - r_j^{-1} z). \tag{36}$$

Since the MA system is assumed to be non-invertible, $|IN1| + |IN2| \geq 1$. From Lemma 1 we have that $\Theta_1(z)\sigma_1^2\Theta_1(\frac{1}{z}) = \Theta_1^\dagger(z)\sigma_1^2\Theta_1^\dagger(\frac{1}{z}) \prod_{r_k \in IN1} |r_k|^{-2}$ and $\Theta_2(z)\sigma_1^2\Theta_2(\frac{1}{z}) = \Theta_2^\dagger(z)\sigma_2^2\Theta_2^\dagger(\frac{1}{z}) \prod_{r_k \in IN2} |r_k|^{-2}$.

Considering the diagonal entries of (34) we note that this implies that

$$A_1(z)\alpha A_1(\frac{1}{z}) = \Phi_1^{-1}(z)\Theta_1^\dagger(z)\Phi_1^{-1}(\frac{1}{z})\Theta_1^\dagger(\frac{1}{z})\sigma_1^2 \prod_{r_k \in IN1} |r_k|^{-2}$$

and

$$A_2(z)\gamma A_2(\frac{1}{z}) = \Phi_2^{-1}(z)\Theta_2^\dagger(z)\Phi_2^{-1}(\frac{1}{z})\Theta_2^\dagger(\frac{1}{z})\sigma_2^2 \prod_{r_k \in IN2} |r_k|^{-2}.$$

Since $A_1(z)$ and $A_2(z)$ must have a leading coefficient of 1, it must be that $A_1(z) = \Phi_1^{-1}(z)\Theta_1^\dagger(z)$ and $A_2(z) = \Phi_2^{-1}(z)\Theta_2^\dagger(z)$.

Now considering the non-diagonal entries of (34) and cross multiplying we obtain

$$\frac{\Phi_1^{-1}(z)\Theta_1(z)\Theta_2(\frac{1}{z})\Phi_2^{-1}(\frac{1}{z})}{A_1(z)A_2(\frac{1}{z})} = \frac{\beta}{\sigma_{12}} = \frac{\Phi_2^{-1}(z)\Theta_2(z)\Theta_1(\frac{1}{z})\Phi_1^{-1}(\frac{1}{z})}{A_2(z)A_1(\frac{1}{z})}. \tag{37}$$

It follows that

$$\frac{\Theta_1(z)\Theta_2(\frac{1}{z})}{\Theta_1^\dagger(z)\Theta_2^\dagger(\frac{1}{z})} = \frac{\Theta_2(z)\Theta_1(\frac{1}{z})}{\Theta_2^\dagger(z)\Theta_1^\dagger(\frac{1}{z})}. \tag{38}$$

Let $\{r_{1,k}\}$ and $\{r_{2,k}$ be roots of $\Theta_1(z)$ and $\Theta_2(z)$ and $OUT1$ and $OUT2$ be the list of roots of $\Theta_1(z)$ and $\Theta_2(z)$ which are outside the unit circle. We can rewrite (38) as

$$\frac{\prod_{r_{1,k}\in IN1}(1-\frac{z}{r_{1,k}})\prod_{r_{1,k}\in OUT1}(1-\frac{z}{r_{1,k}})\prod_{r_{2,k}\in IN2}(1-\frac{1}{r_{2,k}z})\prod_{r_{2,k}\in OUT2}(1-\frac{1}{r_{2,k}z})}{\prod_{r_{1,k}\in IN1}(1-zr_{1,k})\prod_{r_{1,k}\in OUT1}(1-\frac{z}{r_{1,k}})\prod_{r_{2,k}\in IN2}(1-\frac{r_{2,k}}{z})\prod_{r_{2,k}\in OUT2}(1-\frac{1}{r_{2,k}z})} = \frac{\prod_{r_{2,k}\in IN2}(1-\frac{z}{r_{2,k}})\prod_{r_{2,k}\in OUT2}(1-\frac{z}{r_{2,k}})\prod_{r_{1,k}\in IN1}(1-\frac{1}{r_{1,k}z})\prod_{r_{1,k}\in OUT1}(1-\frac{1}{r_{1,k}z})}{\prod_{r_{2,k}\in IN2}(1-zr_{2,k})\prod_{r_{2,k}\in OUT2}(1-\frac{z}{r_{2,k}})\prod_{r_{1,k}\in IN1}(1-\frac{r_{1,k}}{z})\prod_{r_{1,k}\in OUT1}(1-\frac{1}{r_{1,k}z})}$$

or equivalently,

$$\frac{\prod_{r_{1,k}\in IN1}(1-\frac{z}{r_{1,k}})\prod_{r_{2,k}\in IN2}(1-\frac{1}{r_{2,k}z})}{\prod_{r_{1,k}\in IN1}(1-zr_{1,k})\prod_{r_{2,k}\in IN2}(1-\frac{r_{2,k}}{z})} = \frac{\prod_{r_{2,k}\in IN2}(1-\frac{z}{r_{2,k}})\prod_{r_{1,k}\in IN1}(1-\frac{1}{r_{1,k}z})}{\prod_{r_{2,k}\in IN2}(1-zr_{2,k})\prod_{r_{1,k}\in IN1}(1-\frac{r_{1,k}}{z})} \tag{39}$$

where equality holds if and only if the roots in $IN1$ including their multiplicities are identical to those in $IN2$. Thus, if $IN1 \neq IN2$, then equation (38) does not hold and we have a contradiction. Thus the theorem is proven. □

The Wold representation of $\{(X_t, Y_t)'\}$ is crucial in obtaining best linear forecasts under process sharing, obtaining their MSFEs and in determining whether process sharing is valuable (see (15)). Theorem 3 indicates many instances where the Wold representation of $\{(X_t, Y_t)'\}$ is a bivariate model which cannot be obtained from considering the univariate models separately. Therefore it is not sufficient to consider the retailer's order process as two univariate systems as in (10) and (11).

Tunicliffe-Wilson [10] provides an iterative algorithm for determining the Wold representation for the bivariate MA(q) model (29) where $\{(X_t, Y_t)'\}$ is not invertible with respect to $(\epsilon_{1,t}, \epsilon_{2,t})'$.

The algorithm also provides the covariance matrix of the Wold shocks. The following remark is obtained from the discussion surrounding Equation (3.4) of the aforementioned paper.

Remark 4 *The Wold representation of* $\{(X_t, Y_t)'\}$ *is given by*

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \Theta^{\ddagger}(B) \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}, \tag{40}$$

with $\Theta^{\ddagger}(z) = M_0 + M_1 z + .. + M_q z^q$ *where* $M_k = A_k A_0^{-1}$ *such that* A_k *are the converged quantities of* $A_{\tau,k}$ *as obtained from the following iterative system:*

$$\sum_{j=0}^{q-k} (A_{\tau+1,j+k} A'_{\tau,j} + A_{\tau,j+k} A'_{\tau+1,j}) = \Gamma_k + \sum_{j=0}^{q-k} A_{\tau,j+k} A'_{\tau,j}. \tag{41}$$

with $A_{\tau,0}$ *constrained to be upper-triangular and* $\Gamma_k = E[(X_t, Y_t)'(X_{t+k}, Y_{t+k})]$. *Furthermore* $cov(e_{1,t}, e_{2,t})' = A_0 A_0'$. *Convergence of (41) is guaranteed as long as* $\sum_{k=0}^{q} A_{0,k} z^k$ *is nonsingular for* $|z| \leq 1$. *We note from the definition of* $\Theta^{\ddagger}(z)$ *that the Wold representation is MA(q).*

In the examples below we utilize the above remark to obtain the Wold representation and determine the MSFE of the best linear forecast one-step ahead. We also use Equation (8) of Kohn [5] to determine whether process sharing is valuable. Theorem 4, Corollary 2 and Theorem 5 of Section 4 are used to compute MSFEs under shock sharing, process sharing and no sharing respectively.

Example 4 *Suppose the retailer processes can be modeled as MA(1) given by*

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - .5B & 0 \\ 0 & 1 - 2B \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{42}$$

where $cov[(\epsilon_{1,t}, \epsilon_{2,t})'] = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$.

We note that $\{(X_t, Y_t)'\}$ is not invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ and the MA matrix in the Wold representation is non-diagonal. Computing the MSFE of the one-step ahead forecast under no sharing yields 7.5322. Furthermore we observe that the MSFE under shock sharing is given by $1+1+2\cdot.5=3$. In order to compute the MSFE under process sharing we use the Tunicliffe-Wilson algorithm to determine the Wold representation, which is given by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1-.5B & 0 \\ -.75B & 1-2B \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}, \tag{43}$$

where $cov[(e_{1,t}, e_{2,t})'] = \begin{pmatrix} 1 & .5 \\ .5 & 3.25 \end{pmatrix}$. Thus the corresponding MSFE of the one-step ahead forecast under process sharing is $1+3.25+2\cdot.5 = 5.25$ as per Theorem 4 since $\{e_{1,n}\}_{n=-\infty}^{t}$ and $\{e_{2,n}\}_{n=-\infty}^{t}$ are observable. We observe that this is significantly lower than the MSFE under no sharing and significantly higher than the MSFE under shock sharing.

In the following example, we compare the MSFEs under the three sharing arrangements for different values of the $\theta_{2,1}$ coefficient. This is performed in the following example by determining the Wold representation and using Theorem 4 to find the MSFE under process sharing.

Example 5 *Consider the family of models given by*

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1-.5B & 0 \\ 0 & 1-\theta_{2,1}B \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \tag{44}$$

where $cov[(\epsilon_{1,t}, \epsilon_{2,t})'] = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$ *and* $-5 < \theta_{2,1} < 5$.

Varying $\theta_{2,1}$ between -5 and 5 and computing the ratios of the MSFE under no sharing to the MSFE under process sharing as well as the ratios of the MSFE under no sharing to the MSFE under shock sharing we obtain Figure 3. We note that there are *three* locations where process sharing has no value. As should be expected, process sharing and shock sharing are equivalent in the invertibility region.

We note that two of the locations ($\theta_{2,1} = -2.30277$ and $\theta_{2,1} = 1.30279$) are outside the invertibility region, and will result in a non-diagonal MA matrix in the Wold representation as discussed. We can show that these two locations correspond to condition (8) in Theorem 1 of Kohn [5] . For instance, if $\theta_{2,1} = -2.30277$ the Wold representation is given by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - .5B & 0 \\ .9342552B & 1 + .43425961B \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}. \tag{45}$$

with $A(1) = \begin{pmatrix} -.5 & 0 \\ .93426 & .43426 \end{pmatrix}$ in (15) and it is easy to verify that $A(1)'(1,1)' = .432426(1,1)'$. The same could be observed if $\theta_{2,1} = 1.30279$.

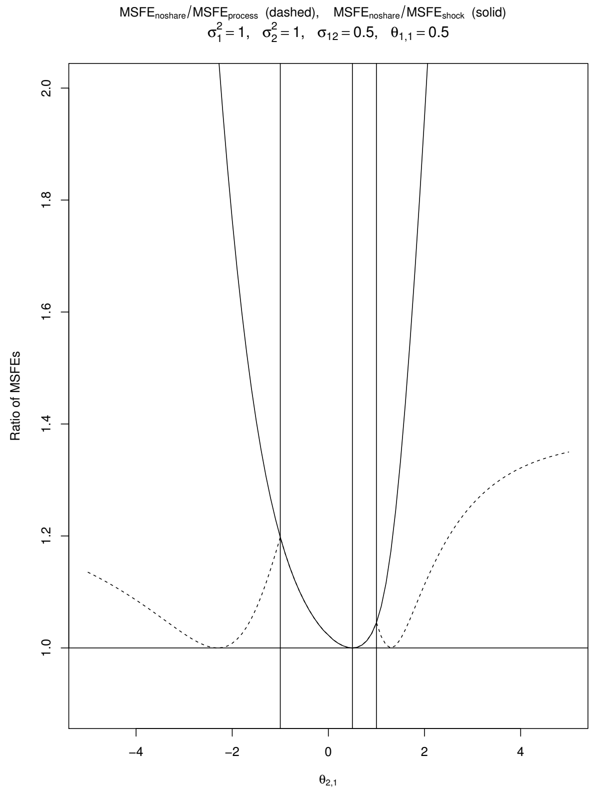

Figure 3: The one-step ahead MSFE ratio of no sharing to process sharing and MSFE ratio of no sharing to shock sharing is shown for the parameters in (44). Vertical lines are drawn at -1 and 1 to indicate the region of invertibility which lies between them. We include an additional vertical line at the location where no sharing, process sharing, and shock sharing result in the same MSFE. A horizontal line is drawn at 1 to indicate instances of no value to process sharing.

As indicated in Figure 3, for any value $\theta_{1,1}$ there are three values of $\theta_{2,1}$ where there is no value to process sharing. This is shown succinctly in Figure 4, where for each $\theta_{1,1}$, a grid search is carried out for values of $\theta_{2,1}$ such that there is no value to process sharing. Each dot represents a pair of values for $\theta_{1,1}$ and $\theta_{2,1}$ such that there is no value to process sharing.

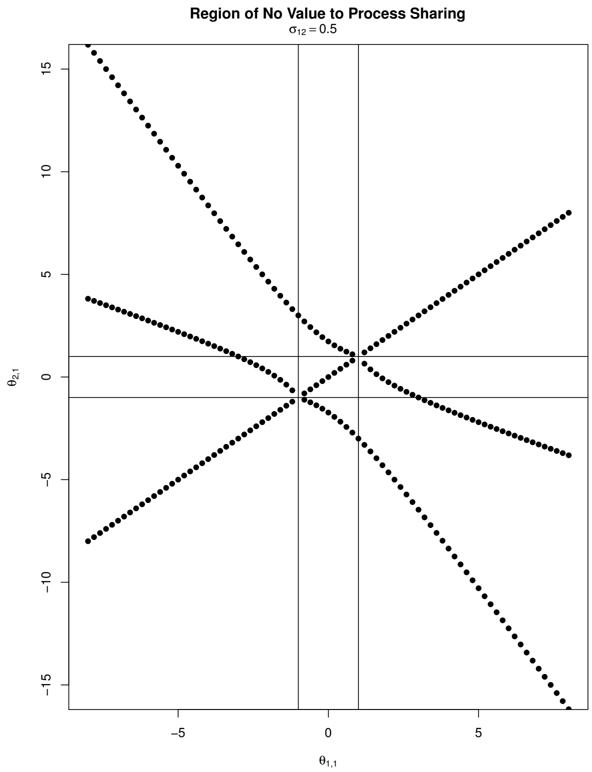

Figure 4: Region of no value to process sharing. Each dot represents a pair of coefficients $\theta_{1,1}$ and $\theta_{2,1}$ where the forecasts under process sharing and no sharing are the same.

4 Computing Mean Squared Forecast Errors

In this section we discuss how to compute MSFEs for the best linear forecasts of Z_{t+1} at time t under shock sharing, process sharing, and no sharing for the various cases as discussed in Section 3.

The next theorem provides the MSFE under the assumptions that the shock sequences are observable to the supplier.

Theorem 4 *Suppose the retailer's processes can be modeled as in (12). If* $\{\epsilon_{1,n}\}_{n=-\infty}^{t}$ *and* $\{\epsilon_{2,n}\}_{n=-\infty}^{t}$

are observable then the best linear forecast of Z_{t+1} is given by

$$(1,1)(\Psi(B)-I)\begin{pmatrix}\epsilon_{1,t}\\ \epsilon_{2,t}\end{pmatrix} \tag{46}$$

where I is a 2×2 identity matrix and $\Psi(B)=\begin{pmatrix}\frac{\Theta_1(B)}{\Phi_1(B)} & 0\\ 0 & \frac{\Theta_2(B)}{\Phi_2(B)}\end{pmatrix}$. *Its MSFE is $\sigma_1^2+\sigma_2^2+2\cdot\sigma_{12}$, which is the sum of the elements of the covariance matrix of $(\epsilon_{1,t},\epsilon_{2,t})'$.*

The Proof of Theorem 4 can be found in the Appendix. Under shock sharing, the retailer has provided the supplier with sequences $\{\epsilon_{1,n}\}_{n=-\infty}^{t}$ and $\{\epsilon_{2,n}\}_{n=-\infty}^{t}$. Hence the best linear forecast and the MSFE under shock sharing is provided by Theorem 4. There is another situation under which the shock sequences are observable to the supplier. This occurs if the retailer shares the processes $\{X_t\}$ and $\{Y_t\}$, and the bivariate process is invertible with respect to $\{(\epsilon_{1,t},\epsilon_{2,t})'\}$ in (12). This leads to the following corollary.

Corollary 1 *Under process sharing, if det* $\begin{pmatrix}\Theta_1(z) & 0\\ 0 & \Theta_2(z)\end{pmatrix}$ *has no roots on or inside the unit circle then the best linear forecast of Z_{t+1} and its MSFE are given by Theorem 4.*

If *det* $\begin{pmatrix}\Theta_1(z) & 0\\ 0 & \Theta_2(z)\end{pmatrix}$ does have a root inside the unit circle however, $\{(\epsilon_{1,n},\epsilon_{2,n})'\}_{n=-\infty}^{t}$ are not observable to the supplier, even when the retailer shares $\{X_n\}_{n=-\infty}^{t}$ and $\{Y_n\}_{n=-\infty}^{t}$. In order to obtain one-step ahead forecasts and MSFEs in this case, we must first obtain the Wold representation of $\{(X_t,Y_t)'\}$ with respect to the Wold shocks $\{(e_{1,t},e_{2,t})'\}$, or equivalently the ARMA representation of $\{(X_t,Y_t)'\}$ with respect to $\{(e_{1,t},e_{2,t})'\}$ (refer to Sections 3.2 and 3.3). As highlighted by Proposition 1 and Theorem 3, the Wold shocks and resulting ARMA representation will be different depending on whether or not the shocks $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$ are contemporaneously correlated. The following corollary describes the uncorrelated case.

Corollary 2 *Under process sharing, if det* $\begin{pmatrix} \Theta_1(z) & 0 \\ 0 & \Theta_2(z) \end{pmatrix}$ *does have a root inside the unit circle and* $\sigma_{12} = 0$*, then the best linear forecast of* Z_{t+1} *is*

$$(1,1)(\Psi(B) - I)\begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} \tag{47}$$

where I *is a* 2×2 *identity matrix and* $\Psi(B) = \begin{pmatrix} \frac{\Theta_1^\dagger(B)}{\Phi_1(B)} & 0 \\ 0 & \frac{\Theta_2^\dagger(B)}{\Phi_2(B)} \end{pmatrix}$*. Its MSFE is* $\sigma_1^2 \prod_{r_j \in IN1} |r_j|^{-2} + \sigma_2^2 \prod_{r_j \in IN2} |r_j|^{-2}$ *where* $IN1$ *(and* $IN2$*) is the list of all roots,* r_k*, of* $\Theta_1(z)$ *(and* $\Theta_2(z)$*) with repeated entries to allow for multiplicities, such that* $|r_k| < 1$*.*

The proof of Corollary 2 can be found in the Appendix. As discussed in Section 3.3, if $\sigma_{12} \neq 0$, the polynomials in the MA matrix appearing in the Wold representation are different from the polynomials we would observe when inverting the individual univariate processes. The following corollary discusses how to compute the forecast and MSFE when $\sigma_{12} \neq 0$ when the retailer's processes can be modeled as a bivariate MA(q), as in (29).

Corollary 3 *Under process sharing, if the retailer's processes generated by model (29) where det* $\begin{pmatrix} \Theta_1(z) & 0 \\ 0 & \Theta_2(z) \end{pmatrix}$ *does have a root inside the unit circle and* $\sigma_{12} \neq 0$*, then the best linear forecast of* Z_{t+1} *is*

$$(1,1)(\Theta^\ddagger(B) - I)\begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} \tag{48}$$

where I *is a* 2×2 *identity matrix. Its MSFE is given by the sum of the elements of* $cov(e_{1,t}, e_{2,t})'$*.*

We note that both $\Theta^\ddagger$ and $cov(e_{1,t}, e_{2,t})'$ can be obtained using Remark 4. The proof of Corollary 3 can be found in the Appendix.

We next consider the no sharing case and forecast Z_{t+1} based on $\{Z_n\}_{n=-\infty}^{t}$. Theorem 5 describes how to obtain the variance of the shocks appearing in the Wold representation of Z_t in this case, which is equivalent to the one-step ahead MSFE. We note that this material is stated without proof in Lemma 1 of Sayed and Kailath [9] . For the sake of completeness and to keep the material in this paper self contained we provide the proof, along with some additional lemmas in the Appendix.

Theorem 5 *The variance of the shocks appearing in the Wold representation of $\{Z_t\}$ is given by*

$$\sigma_{\epsilon_z}^2 = \frac{p_m \prod_{j=1}^{m}(-a_j)}{q_n \prod_{j=1}^{n}(-b_j)} \tag{49}$$

where $\{a_j\}$ and $\{b_j\}$ are the roots of $P(z)$ and $Q(z)$ which are outside the unit circle and p_m is the coefficient of z^m in $P(z)$ and q_n is the coefficient of z^n in $Q(z)$ such that the covariance matrix generating function $S_Z(z)$ of $\{Z_t\}$ can be expressed as the ratio $\frac{O(z)P(z)}{Q(z)}$ where $O(z)$, $P(z)$ and $Q(z)$ are Laurent polynomials[4], with $O(z)$ having all its roots on the unit circle and $P(z)$ and $Q(z)$ having no roots on the unit circle .

We provide technical details on how to obtain the polynomials $O(z)$, $P(z)$ and $Q(z)$ in the Appendix. The variance $\sigma_{\epsilon_z}^2$ of the shocks appearing in the Wold representation of $\{Z_t\}$ is equivalent to the one-step ahead MSFE under no sharing. This exhausts the various MSFE computations surrounding the different sharing arrangements discussed in this paper.

5 Conclusion

In this paper, we consider the case that the retailer observes demand for an item from two different demand streams. We show how this demand propagates to two separate order processes and study

[4]See Definition 2 in the Appendix

whether there is value to sharing these individual processes with a supplier in a two-stage supply chain. We also study whether there is additional value to sharing the retailer's shocks. In order to do so, we study the retailer's order components as part of a bivariate ARMA process. The contribution of our research is determining how to assess whether or not information sharing is valuable when the retailer's order is composed of aggregate processes.

Indeed, we have uncovered situations where there is no value to information sharing even though the existing literature suggested otherwise. Our framework and results are important for future supply chain research when there are multiple processes incorporated into the retailer's order to the supplier.

References

[1] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, 2nd edition, 1991.

[2] Ruomeng Cui, Gad Allon, Achal Bassamboo, and Jan A. Van Mieghem. Information sharing in supply chains: An empirical and theoretical valuation. *Management Science*, 61(11):2803–2824, November 2015.

[3] Vishal Gaur, Avi Giloni, and Sridhar Seshadri. Information Sharing in a Supply Chain Under ARMA Demand. *Management Science*, 51(6):961–969, June 2005.

[4] Avi Giloni, Clifford Hurvich, and Sridhar Seshadri. Forecasting and information sharing in supply chains under arma demand. *IIE Transactions*, 46(1):35–54, January 2014.

[5] Robert Kohn. When is an aggregate of a time series efficiently forecast by its past. *Journal of Econometrics*, 18(11):337–349, April 1982.

[6] Vladimir Kovtun, Avi Giloni, and Clifford Hurvich. Assessing the value of demand sharing in supply chains. *Naval Research Logistics*, 61(7):515–531, October 2014.

[7] Hau L. Lee, Kut C. So, and Christopher S. Tang. The Value of Information Sharing in a Two-Level Supply Chain. *Management Science*, 46(5):626–643, May 2000.

[8] Srinivasan Raghunathan. Information Sharing in a Supply Chain: A Note on its Value when Demand Is Nonstationary. *Management Science*, 47(4):605–610, April 2001.

[9] Ali H. Sayed and Thomas Kailath. A survey of spectral factorization methods. *Numerical Linear Algebra with Applications*, 8(6):467–496, August 2001.

[10] Granville Tunicliffe-Wilson. The factorization of matricial spectral densities. *SIAM Journal of Applied Mathematics*, 23(4):420–426, December 1972.

[11] Xiaolong Zhang. Technical Note: Evolution of ARMA Demand in Supply Chains. *Manufacturing & Service Operations Management*, 6(2):195–198, 2004.

Appendix

Proof of Theorem 1: Since we are assuming that $\{(X_t, Y_t)'\}$ is invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, we note that $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ are the Wold shocks and the Wold representation is given by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \Psi_1(B) & 0 \\ 0 & \Psi_2(B) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \tag{50}$$

where $\Psi_1(z) = \dfrac{\Theta_1(z)}{\Phi_1(z)} = 1 + \Psi_{1,1}z + \Psi_{1,2}z^2 + \ldots$ and $\Psi_2(z) = \dfrac{\Theta_2(z)}{\Phi_2(z)} = 1 + \Psi_{2,1}z + \Psi_{2,2}z^2 + \ldots$. Theorem 1 of Kohn states that there is no value to observation sharing (and hence also shock sharing in this case) if and only if the condition in (15) holds. We have $A(j) = \begin{pmatrix} \Psi_{1,j} & 0 \\ 0 & \Psi_{2,j} \end{pmatrix}$.

Hence, Condition (15) becomes equivalent to

$$\begin{pmatrix} \Psi_{1,j} \\ \Psi_{2,j} \end{pmatrix} = \begin{pmatrix} k_j \\ k_j \end{pmatrix}, j = 1, 2, \ldots \tag{51}$$

or equivalently, $\Psi_{1,j} = \Psi_{2,j}$ for all j. This occurs if and only if $\frac{\Theta_1(z)}{\Phi_1(z)} = \frac{\Theta_2(z)}{\Phi_2(z)}$. $\square$

The following lemma provides a useful factorization of the product of $\Theta(z)\Theta(1/z)$ using the polynomial $\Theta^\dagger(z)$. As will be discussed in the proof of Proposition 1, this factorization is helpful in determining the ARMA representation of $\{(X_t, Y_t)'\}$ with respect to their Wold shocks.

Lemma 1 *For any polynomial $\Theta(z)$ with leading coefficient 1 and no roots on the unit circle,*

$$\Theta(z)\Theta(1/z) = \Theta^\dagger(z)\Theta^\dagger(1/z) \prod_{r_k \in IN} |r_k|^{-2}$$

where IN is the list of all roots, r_k, of $\Theta(z)$ with repeated entries to allow for multiplicities, such that $|r_k| < 1$.

Proof of Lemma 1: Let $\Theta(z)$ be of degree q. For any $z_0 \neq 0$,

$$\left(1 - \frac{z}{z_0}\right)\left(1 - \frac{1}{z\bar{z}_0}\right) = (1 - \bar{z}_0 z)(1 - z_0/z)|z_0|^{-2}.$$

where $\bar{z}_0$ is the conjugate of z_0. Let IN be the list of roots of Θ inside the unit circle and OUT be the list of roots of Θ outside the unit circle and using this fact in the third line below, we obtain

$$\Theta(z)\Theta(1/z) = \Pi_{k=1}^{q}(1 - z/r_k)\left(1 - \frac{1}{zr_k}\right)$$
$$= \prod_{r_k \in IN}(1 - z/r_k)\left(1 - \frac{1}{z\bar{r}_k}\right) \prod_{r_k \in OUT}(1 - z/r_k)\left(1 - \frac{1}{z\bar{r}_k}\right)$$

$$= \prod_{r_k \in IN} (1 - \bar{r}_k z)(1 - r_k/z)|r_k|^{-2} \prod_{r_k \in OUT} (1 - z/r_k) \left(1 - \frac{1}{z\bar{r}_k}\right)$$

$$= \prod_{r_k \in IN} (1 - r_k z)(1 - r_k/z)|r_k|^{-2} \prod_{r_k \in OUT} (1 - z/r_k) \left(1 - \frac{1}{z r_k}\right)$$

$$= \Theta^\dagger(z)\Theta^\dagger(1/z) \prod_{r_k \in IN} |r_k|^{-2}. \square$$

The proof of Proposition 1 also requires the use of a covariance matrix generating function which we describe below. Consider the lag-h covariance matrix $\Gamma(h) = E[(X_{t+h}, Y_{t+h})'(X_t, Y_t)]$. The covariance matrix generating function of $(X_t, Y_t)'$ is $G(z) = \sum_{-\infty}^{\infty} \Gamma(h) z^h$. The covariance matrix generating function of the bivariate ARMA $\{(X_t, Y_t)'\}$ with shocks $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ is given by (see Brockwell and Davis p 420 [1] Equation (11.3.17))

$$G(z) = \begin{pmatrix} \Phi_1(z) & 0 \\ 0 & \Phi_2(z) \end{pmatrix}^{-1} \begin{pmatrix} \Theta_1(z) & 0 \\ 0 & \Theta_2(z) \end{pmatrix} \sum\nolimits_\epsilon \begin{pmatrix} \Theta_1(\frac{1}{z}) & 0 \\ 0 & \Theta_2(\frac{1}{z}) \end{pmatrix} \begin{pmatrix} \Phi_1(\frac{1}{z}) & 0 \\ 0 & \Phi_2(\frac{1}{z}) \end{pmatrix}^{-1} \tag{52}$$

Proof of Proposition 1: Since it is assumed that the bivariate process in (12) is not invertible with respect to the given shocks, at least one of $\{X_t\}$ or $\{Y_t\}$ is not invertible with respect to $\{\epsilon_{1,t}\}$ or $\{\epsilon_{2,t}\}$ respectively.

Define $\{e_{1,t}, e_{2,t}\}$ as

$$\begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} = \begin{pmatrix} \dfrac{\Phi_1(B)}{\Theta_1^\dagger(B)} & 0 \\ 0 & \dfrac{\Phi_2(B)}{\Theta_2^\dagger(B)} \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix}. \tag{53}$$

We first note that $(e_{1,t}, e_{2,t})'$ is a linear combination of elements in $\{(X_n, Y_n)'\}_{n=-\infty}^t$ since $\Theta_1^\dagger(z)$ and $\Theta_2^\dagger(z)$ have no roots inside the unit circle by definition. Furthermore $(X_t, Y_t)'$ is a linear combination of elements in $\{(e_{1,n}, e_{2,n})'\}_{n=-\infty}^t$, which can be seen rewriting (53) as

$$\begin{pmatrix} \dfrac{\Theta_1^\dagger(B)}{\Phi_1(B)} & 0 \\ 0 & \dfrac{\Theta_2^\dagger(B)}{\Phi_2(B)} \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} = \begin{pmatrix} X_t \\ Y_t \end{pmatrix} \tag{54}$$

and recalling that $\Phi_1(z)$ and $\Phi_2(z)$ have all their roots outside the unit circle. Also from (53) we observe that

$$\begin{pmatrix} \Theta_1^\dagger(B) & 0 \\ 0 & \Theta_2^\dagger(B) \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} = \begin{pmatrix} \Phi_1(B) & 0 \\ 0 & \Phi_2(B) \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix}. \tag{55}$$

It remains to show that $\{(e_{1,t}, e_{2,t})'\}$ is a bivariate white noise sequence. Comparing (55) with (12) we observe that

$$\begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} = \begin{pmatrix} \frac{\Theta_1(B)}{\Theta_1^\dagger(B)} & 0 \\ 0 & \frac{\Theta_2(B)}{\Theta_2^\dagger(B)} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \tag{56}$$

Based on this and Σ_ϵ, the covariance matrix generating function (see (52) of $\{(e_{1,t}, e_{2,t})'\}$ is given by

$$\begin{pmatrix} \frac{\Theta_1(z)}{\Theta_1^\dagger(z)} & 0 \\ 0 & \frac{\Theta_2(z)}{\Theta_2^\dagger(z)} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \frac{\Theta_1(1/z)}{\Theta_1^\dagger(1/z)} & 0 \\ 0 & \frac{\Theta_2(1/z)}{\Theta_2^\dagger(1/z)} \end{pmatrix} \tag{57}$$

which can be rewritten as

$$\begin{pmatrix} \frac{\sigma_1^2\Theta_1(z)\Theta_1(1/z)}{\Theta_1^\dagger(z)\Theta_1^\dagger(1/z)} & 0 \\ 0 & \frac{\sigma_2^2\Theta_2(z)\Theta_2(1/z)}{\Theta_2^\dagger(z)\Theta_2^\dagger(1/z)} \end{pmatrix}. \tag{58}$$

From the factorization in Lemma 1 we can rewrite this covariance matrix generating function as

$$\begin{pmatrix} \sigma_1^2 \prod_{r_j \in IN1} |r_j|^{-2} & 0 \\ 0 & \sigma_2^2 \prod_{r_j \in IN2} |r_j|^{-2} \end{pmatrix}. \tag{59}$$

This shows that $\{(e_{1,t}, e_{2,t})'\}$ is a bivariate white noise sequence with covariance matrix

$$\sum_e = \begin{pmatrix} \sigma_1^2 \prod_{r_j \in IN1} |r_j|^{-2} & 0 \\ 0 & \sigma_2^2 \prod_{r_j \in IN2} |r_j|^{-2} \end{pmatrix}.$$

□

Proof of Theorem 4: Let $\Phi(z)$ and $\Theta(z)$ be such that $\Phi(B) = \begin{pmatrix} \Phi_1(B) & 0 \\ 0 & \Phi_2(B) \end{pmatrix}$ and $\Theta(B) = \begin{pmatrix} \Theta_1(B) & 0 \\ 0 & \Theta_2(B) \end{pmatrix}$ in (12). We can then rewrite (12) as

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \Phi^{-1}(B)\Theta(B) \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{60}$$

which is an MA(∞) representation of $\{(X_t, Y_t)'\}$ with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$. $(X_t, Y_t)'$. Furthermore this implies that

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = \Phi^{-1}(B)\Theta(B) \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} \tag{61}$$

or equivalently

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} + (\Phi^{-1}(B)\Theta(B) - I) \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}. \tag{62}$$

where I is a 2×2 identity matrix. Since we are interested in the forecast of $Z_{t+1} = X_{t+1} + Y_{t+1} = (1,1)(X_{t+1}, Y_{t+1})'$, we rewrite (62) as

$$(1,1)\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = (1,1)\begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} + (1,1)(\Phi^{-1}(B)\Theta(B) - I) \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} \tag{63}$$

We note that $(1,1)(\Phi^{-1}(B)\Theta(B) - I) \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}$ includes only elements in the series $\{(\epsilon_{1,n}, \epsilon_{2,n})'\}_{n=-\infty}^{t}$, which are all observable by assumption, and therefore the best linear forecast of $X_{t+1} + Y_{t+1}$ is given by

$$(1,1)(\Psi(B) - I) \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}. \tag{64}$$

To get the MSFE of this best linear forecast note that

$$\begin{aligned} MSFE &= E[(X_{t+1} + Y_{t+1} - (1,1)(\Psi(B) - I)(\epsilon_{1,t}, \epsilon_{2,t})')^2] &(65)\\ &= E[(\epsilon_{1,t+1} + \epsilon_{2,t+1})^2] &(66)\\ &= E[\epsilon_{1,t+1}^2 + \epsilon_{2,t+1}^2 + 2 \cdot \epsilon_{1,t+1}\epsilon_{2,t+1}] &(67)\\ &= \sigma_1^2 + \sigma_2^2 + 2 \cdot \sigma_{12} &(68)\end{aligned}$$

□

Proof of Theorem 2: From Proposition 1, if $\sigma_{12} = 0$ then the ARMA representation of $\{(X_t, Y_t)'\}$ with respect to $\{(e_{1,t}, e_{2,t})'\}$ is given by

$$\Phi(B) \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \Theta^{\dagger}(B) \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} \tag{69}$$

where $\Phi(B) = \begin{pmatrix} \Phi_1(B) & 0 \\ 0 & \Phi_2(B) \end{pmatrix}$ and $\Theta^{\dagger}(B) = \begin{pmatrix} \Theta_1^{\dagger}(B) & 0 \\ 0 & \Theta_2^{\dagger}(B) \end{pmatrix}$, with $cov[(e_{1,t}, e_{2,t})'] = \begin{pmatrix} \sigma_1^2 \prod_{r_j \in IN1} |r_j|^{-2} & 0 \\ 0 & \sigma_2^2 \prod_{r_j \in IN2} |r_j|^{-2} \end{pmatrix}$.

Since $det(\Theta^{\dagger}(z))$ has no roots inside the unit circle, from Corollary 1 we obtain the desired result. □

Proof of Corollary 3: If the retailer observes the bivariate MA(q) model in (29) such that $\{(X_t, Y_t)'\}$ is noninvertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ and $\sigma_{12} \neq 0$ then from Remark 4 we obtain the Wold representation given by:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \Theta^{\ddagger}(B) \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} \tag{70}$$

where $det(\Theta^{\ddagger}(z))$ has no roots inside the unit circle. From this we observe that

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = \Theta^{\ddagger}(B) \begin{pmatrix} e_{1,t+1} \\ e_{2,t+1} \end{pmatrix} \tag{71}$$

or equivalently

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = \begin{pmatrix} e_{1,t+1} \\ e_{2,t+1} \end{pmatrix} + (\Theta^{\ddagger}(B) - I) \begin{pmatrix} e_{1,t+1} \\ e_{2,t+1} \end{pmatrix}. \tag{72}$$

where I is a 2×2 identity matrix. Since we are interested in the forecast of $Z_{t+1} = X_{t+1} + Y_{t+1} = (1,1)(X_{t+1}, Y_{t+1})'$, we rewrite (72) as

$$(1,1) \begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = (1,1) \begin{pmatrix} e_{1,t+1} \\ e_{2,t+1} \end{pmatrix} + (1,1)(\Theta^{\ddagger}(B) - I) \begin{pmatrix} e_{1,t+1} \\ e_{2,t+1} \end{pmatrix} \tag{73}$$

We note that $(1,1)(\Theta^{\ddagger}(B)-I) \begin{pmatrix} e_{1,t+1} \\ e_{2,t+1} \end{pmatrix}$ includes only elements in the series $\{(e_{1,n}, e_{2,n})'\}_{n=-\infty}^{t}$, which are all observable by assumption, and therefore the best linear forecast of $X_{t+1} + Y_{t+1}$ is given by

$$(1,1)(\Theta^{\ddagger}(B) - I) \begin{pmatrix} e_{1,t+1} \\ e_{2,t+1} \end{pmatrix}. \tag{74}$$

To get the MSFE of this best linear forecast note that

$$\begin{aligned} MSFE &= E[(X_{t+1} + Y_{t+1} - (1,1)(\Theta^{\ddagger}(B) - I)(e_{1,t}, e_{2,t})')^2] && (75) \\ &= E[(e_{1,t+1} + e_{2,t+1})^2] && (76) \\ &= E[e_{1,t+1}^2 + e_{2,t+1}^2 + 2 \cdot e_{1,t+1}e_{2,t+1}] && (77) \\ &= \text{sum of the elements of } cov(e_{1,t}, e_{2,t})'. && (78) \end{aligned}$$

□

5.1 Examples Illustrating the Value of Information Sharing

We begin with several examples which illustrate Theorem 1, as well as the propagation discussed in Section 2. Methods used to compute the MSFEs under shock sharing and no sharing can be found in Theorem 4 and Theorem 5 of Section 4.

Example 6 *Suppose* $\{D_{1,t}\}$ *and* $\{D_{2,t}\}$ *are each MA(2) with respect to* $\{\tilde{\epsilon}_{1,t}\}$ *and* $\{\tilde{\epsilon}_{2,t}\}$ *respectively, given by*

$$D_{1,t} = (1 - .2B - .4B^2)\tilde{\epsilon}_{1,t} \quad (79)$$

$$D_{2,t} = (1 - .4B - \theta^\star_{2,2}B^2)\tilde{\epsilon}_{2,t}. \quad (80)$$

where $\left|\dfrac{\theta^\star_{2,2}}{1-.4}\right| < 1$. *We note that this restriction guarantees that* $\{D_{2,t}\}$ *is invertible with respect to* $\{\tilde{\epsilon}_{2,t}\}$. *Furthermore this requirement will guarantee the invertibility of* $\{(X_t, Y_t)'\}$ *as well. Using Theorem 3 of KGH, it follows that in this case* $\{(X_t, Y_t)'\}$ *can be represented as the bivariate ARMA process*

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - .5B & 0 \\ 0 & 1 - \dfrac{\theta^\star_{2,2}}{1-.4}B \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \quad (81)$$

where $\left|\dfrac{\theta^\star_{2,2}}{1-.4}\right| < 1$. *Furthermore* $\epsilon_{1,t} = (1-.2)\tilde{\epsilon}_{1,t}$ *and* $\epsilon_{2,t} = (1-.4)\tilde{\epsilon}_{2,t}$. *We note that here* $\theta_{2,1} = .5$ *and* $\theta_{2,2} = \dfrac{\theta^\star_{2,2}}{1-.4}$. *Without loss of generality we assume that* $\sum_\epsilon = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$.

The ratio of MSFEs under no sharing and under information (process or shock) sharing is plotted in Figure 5 for different value of $\theta_{2,2}$ between -.6 and .6.

As Theorem 1 implies (noting that $\Phi_1(z) = \Phi_2(z) = 1$), the MSFEs are equal only when $\theta^\star_{2,2} = .3$ such that $\dfrac{\theta^\star_{2,2}}{1-.4} = \theta_{2,1} = .5 = \theta_{1,1}$. We observe that the ratio of MSFEs is a convex function in $\theta^\star_{2,2}$ reaching its minimum when $\theta_{2,1} = \theta_{1,1}$.

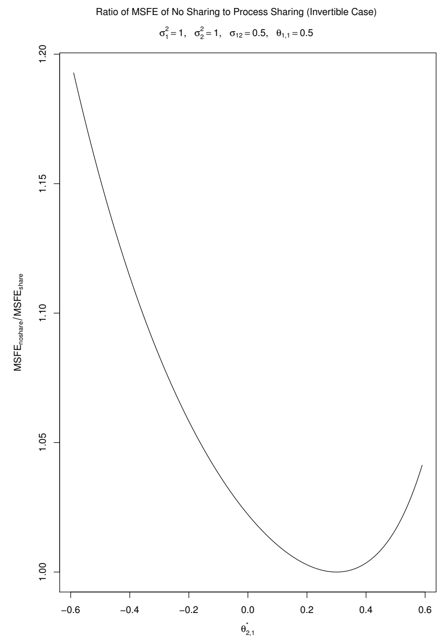

Figure 5: The one-step ahead MSFE ratio of no sharing to information sharing is shown for the parameters in (81).

In order to keep the focus strictly on the value of information sharing, we will forgo discussing propagation in the next example. Instead, we consider the representation of $\{(X_t, Y_t)'\}$ directly[5].

Example 7 *Suppose $\{(X_t, Y_t)'\}$ is bivariate ARMA(1,1) with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ given by*

$$\begin{pmatrix} 1 - .1B & 0 \\ 0 & 1 - .7B \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 - .5B & 0 \\ 0 & 1 - \theta_{2,1}B \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \quad (82)$$

with $\sum_\epsilon = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$ *where* $|\theta_{2,1}| < 1$.

We note that $\{(X_t, Y_t)'\}$ is causal and invertible with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$ for any such $\theta_{2,1}$. The ratio of MSFEs under no sharing and under information (process or shock) sharing is plotted

[5]In general, for any desired model of $\{(X_t, Y_t)'\}$ with respect to $\{(\epsilon_{1,t}, \epsilon_{2,t})'\}$, there exist corresponding model(s) of $\{(D_{1,t}, D_{2,t})'\}$ with respect to $\{(\tilde{\epsilon}_{1,t}, \tilde{\epsilon}_{2,t})'\}$.

in Figure 6 as $\theta_{2,1}$ varies between -1 and 1. A horizontal line is drawn at 1, corresponding to the case that the MSFE under no sharing is equivalent to the MSFE under information sharing. As the theorem implies the MSFEs can never be equal since the entries of the AR matrix are not equal.

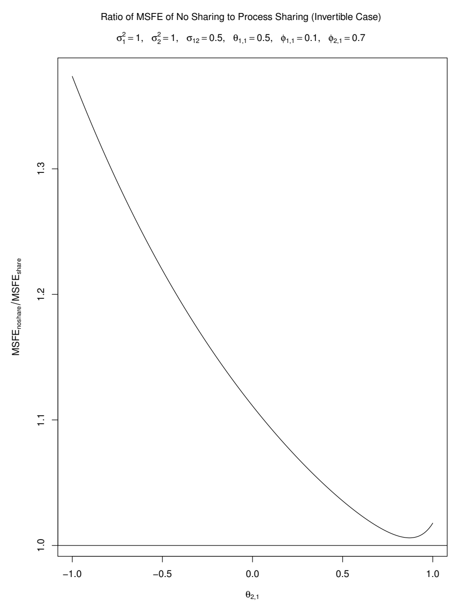

Figure 6: The one-step ahead MSFE ratio of no sharing to information sharing is shown for the parameters in (81)

Obtaining the Variance of the Wold Shocks of Z_t

Here we provide the necessary materials for proving Theorem 5, as well as a construction of the Laurent polynomials $P(z)$, $Q(z)$ and $O(z)$ at the center of the Theorem. We begin with a definition of Laurent polynomials[6] and a development of several lemmas which describe how Laurent polynomials can be factorized. This will be key in obtaining the variance of the shocks appearing in the Wold representation of $\{Z_t\}$.

[6] Although more general definitions of Laurent polynomials exist in the literature, we consider the more restrictive one provided here.

Definition 2 *A function $P(z)$ in a complex variable z is a Laurent polynomial if we can write $P(z)$ as*

$$P(z) = \sum_{k=-m}^{m} p_k z^k$$

where $p_{-m}, \ldots, p_m$ are real-valued coefficients and $p_h = p_{-h}$ for all h. If $p_m \neq 0$ then $P(z)$ is said to be of order m.

Laurent polynomials can be factorized according to the following two Lemmas, which are separated by whether the Laurent polynomial has roots on the unit circle.

Lemma 2 *Let $P(z)$ be a Laurent polynomial of order m with no roots on the unit circle. Then $P(z)$ can be factorized as*

$$P(z) = G(z)G(1/z)p_m \prod_{j=1}^{m}(-r_j) \tag{83}$$

where $G(z)$ is a polynomial with order m, leading coefficient 1, and all roots outside the unit circle with $\{r_j\}$ being the roots of $P(z)$ that are outside the unit circle.

Proof. Suppose z_0 is such that $P(z_0) = 0$. Note that

$$P(1/z_0) = \sum_{k=-m}^{m} p_k z_0^{-k} = \sum_{k=-m}^{m} p_k z_0^{k} = 0$$

and hence $\frac{1}{z_0}$ is also a root of $P(z)$.

Now consider $z^m P(z)$ which is a polynomial of degree $2m$ with a nonzero leading coefficient. Let $r_1, \ldots, r_{2m}$ be its nonzero roots, of which m are outside and m are inside the unit circle. Without loss of generality let $r_1, \ldots, r_m$ be outside the unit circle. It is possible to factorize $z^m P(z)$ as

$$z^m P(z) = p_m \prod_{j=1}^{2m}(1 - z/r_j)$$

where $g_m \neq 0$ and real. Therefore

$$P(z) = p_m \frac{1}{z^m} \prod_{j=1}^{m}(1 - z/r_j)(1 - zr_j)$$

Proceeding further we observe that

$$
\begin{aligned}
P(z) &= p_m \prod_{j=1}^{m} (1 - z/r_j)(\frac{1 - zr_j}{z}) && (84) \\
&= p_m \prod_{j=1}^{m} (1 - z/r_j)(-r_j)(1 - \frac{1}{zr_j}) && (85) \\
&= p_m \prod_{j=1}^{m} (-r_j) \prod_{j=1}^{m} (1 - z/r_j) \prod_{j=1}^{m} (1 - \frac{1}{zr_j}) && (86) \\
& && (87)
\end{aligned}
$$

Now let $G(z) = \prod_{j=1}^{m} (1 - z/r_j)$ and the desired result is achieved. □

Lemma 3 *Let $O(z)$ be a Laurent polynomial of order ON with all its roots on the unit circle such that the coefficient of the highest degree term is 1. Then $O(z)$ can be factorized as*

$$O(z) = R(z)R(1/z)C(z)C(1/z) \tag{88}$$

where any root of $R(z)$ is either 1 or -1 and any root of $C(z)$ is complex and on the unit circle. Furthermore, if $z^{ON}O(z)$ has c complex roots and re real roots, we can list the roots of $C(z)$ as $r_1, \ldots, r_{c/2}$ where each root is above the real axis and list roots of $R(z)$ as $r_{c+1}, \ldots, r_{c+re}$ such that $r_{c+1} = r_{c+re/2+1}, \ldots, r_{c+re/2} = r_{c+re}$.

Proof:

Consider $z^{ON}O(z)$ which is a polynomial of degree $2 \cdot ON$ having all its roots nonzero roots on the unit circle. It is possible to factorize $z^{ON}O(z)$ as

$$z^{ON}O(z) = \prod_{j=1}^{2 \cdot ON} (1 - z/r_j)$$

Without loss of generality let $r_1, \ldots r_c$ be the complex roots of $z^{ON}O(z)$, and $r_{c+1}, \ldots, r_{2 \cdot ON}$ be the re real roots. For each complex root of $z^{ON}O(z)$, we note that the conjugate is also a root and therefore we can consider $r_1, \ldots, r_{c/2}$ to be those complex roots which are above the real axis.

Hence it is possible to factorize $z^{ON}O(z)$ as

$$z^{ON}O(z) = \prod_{j=1}^{c/2}(1 - z/r_j)(1 - zr_j) \prod_{j=c+1}^{c+re}(1 - zr_j).$$

Furthermore we note that any real root (-1 or 1) of $z^{ON}O(z)$ must have even multiplicity. This can be seen by noting that the degree of $z^{ON}O(z)$ is even and therefore it must have an even number of real roots. Furthermore since $z^{ON}O(z)$ and $\prod_{j=1}^{c/2}(1 - z/r_j)(1 - zr_j)$ are Palendromic polynomials, so is $\prod_{j=c+1}^{c+re}(1 - zr_j)$. Since the coefficient of z^0 is 1, this implies that the coefficient of z^{re} must be 1 and therefore the number of roots at -1 must be even. Hence it is possible to relist the real roots such that $r_{c+1} = r_{c+re/2+1}, \ldots, r_{c+re/2} = r_{c+re}$. Thus the previous factorization can be restated as

$$z^{ON}O(z) = \prod_{j=1}^{c/2}(1 - z/r_j)(1 - zr_j) \prod_{j=c+1}^{c+re/2}(1 - z/r_j)(1 - zr_j).$$

Dividing by z^{ON} on both sides we obtain

$$O(z) = \prod_{j=1}^{c/2}(1 - z/r_j)\frac{(1 - zr_j)}{z} \prod_{j=c+1}^{c+re/2}(1 - z/r_j)\frac{(1 - zr_j)}{z}$$

or equivalently

$$O(z) = \prod_{j=1}^{c/2}(1 - z/r_j)(-r_j)(1 - \frac{1}{zr_j}) \prod_{j=c+1}^{c+re/2}(1 - z/r_j)(-r_j)(1 - \frac{1}{zr_j}).$$

Rearranging yields

$$O(z) = \prod_{j=1}^{c/2}(-r_j) \prod_{j=c+1}^{c+re/2}(-r_j) \prod_{j=1}^{c/2}(1 - z/r_j) \prod_{j=1}^{c/2}(1 - \frac{1}{zr_j}) \prod_{j=c+1}^{c+re/2}(1 - z/r_j) \prod_{j=c+1}^{c+re/2}(1 - \frac{1}{zr_j}).$$

Finally we note that $\prod_{j=1}^{c/2}(-r_j) \prod_{j=c+1}^{c+re/2}(-r_j)$ must be equal to 1 since the coefficient of the highest degree term of $O(z)$ is 1 and conclude by letting $C(z) = \prod_{j=1}^{c/2}(1 - z/r_j)$ and $R(z) = \prod_{j=c+1}^{c+re/2}(1 - z/r_j)$.

□

The next lemma shows that we can decompose the covariance matrix generating function of $\{Z_t\}$ into a ratio of Laurent polynomials. This combined with the previous two lemmas will be used to obtain the variance of the shocks appearing in the Wold representation of $\{Z_t\}$.

Lemma 4 *The covariance matrix generating function $S_Z(z)$ of $\{Z_t\}$ can be expressed as the ratio $\frac{O(z)P(z)}{Q(z)}$ where $O(z)$, $P(z)$ and $Q(z)$ are Laurent polynomials, with $O(z)$ having all its roots on the unit circle and $P(z)$ and $Q(z)$ having no roots on the unit circle.*

Proof: Consider the spectral density $f_Z(\lambda)$ of $\{Z_t\}$ which can be shown to be

$$f_Z(\lambda) = f_X(\lambda) + f_Y(\lambda) + f_{XY}(\lambda) + \bar{f}_{XY}(\lambda) \tag{89}$$

where $f_X(\lambda)$ is the spectral density of $\{X_t\}$, $f_Y(\lambda)$ is the spectral density of $\{Y_t\}$, and $f_{XY}(\lambda)$ is the cross-spectrum defined by $f_{XY}(\lambda) = \frac{1}{2\pi}\sum_{r=-\infty}^{\infty} e^{-i\lambda r} c_{XY}(r)$ where $c_{XY}(r) = E[X_{t+r}Y_t]$ is the cross-covariance sequence. For the bivariate ARMA process in (12),

$$f_X(\lambda) = \frac{\sigma_1^2}{2\pi}\frac{|\Theta_1(e^{-i\lambda})|^2}{|\Phi_1(e^{-i\lambda})|^2} \tag{90}$$

$$f_Y(\lambda) = \frac{\sigma_2^2}{2\pi}\frac{|\Theta_2(e^{-i\lambda})|^2}{|\Phi_2(e^{-i\lambda})|^2} \tag{91}$$

$$f_{XY}(\lambda) = \frac{\sigma_{12}}{2\pi}\frac{\Theta_1(e^{-i\lambda})}{\Phi_1(e^{-i\lambda})}\frac{\Theta_2(e^{i\lambda})}{\Phi_2(e^{i\lambda})} \tag{92}$$

$$\bar{f}_{XY}(\lambda) = \frac{\sigma_{12}}{2\pi}\frac{\Theta_1(e^{i\lambda})}{\Phi_1(e^{i\lambda})}\frac{\Theta_2(e^{-i\lambda})}{\Phi_2(e^{-i\lambda})} \tag{93}$$

We note that $S_Z(z) = \sum_{j=-\infty}^{\infty} R_z(j)z^j$ where $R_z(j) = E(Z_t Z_{t-j})$ and therefore we observe the equivalence $S_Z(e^{-i\lambda}) = 2\pi f_Z(\lambda)$. As such, $S_Z(z)$ can be obtained as

$$S_Z(z) = \sigma_1^2\frac{\Theta_1(z)\Theta_1(z^{-1})}{\Phi_1(z)\Phi_1(z^{-1})} + \sigma_2^2\frac{\Theta_2(z)\Theta_2(z^{-1})}{\Phi_2(z)\Phi_2(z^{-1})} + \sigma_{12}\frac{\Theta_1(z)\Theta_2(z^{-1})\Phi_1(z^{-1})\Phi_2(z) + \Theta_1(z^{-1})\Theta_2(z)\Phi_1(z)\Phi_2(z^{-1})}{\Phi_1(z)\Phi_2(z^{-1})\Phi_1(z^{-1})\Phi_2(z)}. \tag{94}$$

Each of the additive terms is a ratio of two Laurent polynomials [7] (defined in Definition 2) and as such $S_Z(z)$ is the ratio of two Laurent polynomials. Without loss of generality, this ratio can be expressed as

$$S_Z(z) = \frac{O(z)P(z)}{Q(z)} \tag{95}$$

where $O(z)$ has all roots on the unit circle with the coefficient of the highest-degree term equal to 1 and $P(z)$, $Q(z)$ have orders m and n respectively and no roots on the unit circle. The fact that the denominator of $S(z)$, which is equal to $\Phi_1(z)\Phi_2(z^{-1})\Phi_1(z^{-1})\Phi_2(z)$ has no roots on the unit circle comes from the fact that $\Phi_1(z)$ and $\Phi_2(z)$ have no roots on the unit circle by assumption $\square$.

We note that (94) and (95) provide the construction of the Laurent polynomials $O(z)$, $P(z)$ and $Q(z)$ in Theorem 5. This, along with their factorizations, allows us to establish the variance of the shocks appearing in the Wold representation of $\{Z_t\}$ in the proof of Theorem 5 below.

Proof of Theorem 5: By Lemma 4 we observe that the covariance matrix generating function $S_Z(z)$ of $\{Z_t\}$ can be expressed as the ratio $\dfrac{O(z)P(z)}{Q(z)}$ where $P(z)$ and $Q(z)$ have order m and n and roots $\{a_j\}$ and $\{b_j\}$ which are outside the unit circle. Using Lemma 2, we can factorize $P(z)$ and $Q(z)$ as

$$P(z) \;=\; p_m \prod_{j=1}^{m}(-a_j)L_p(z)L_p(\frac{1}{z}) \tag{96}$$

$$Q(z) \;=\; q_n \prod_{j=1}^{n}(-b_j)L_q(z)L_q(\frac{1}{z}) \tag{97}$$

such that $L_p(z)$ and $L_q(z)$ have leading coefficient 1 and roots $\{a_j\}$ and $\{b_j\}$ respectively. From the proof of Lemma 2 we observe that

$$L_p(z) = \prod_{j=1}^{m}(1 - z/a_j)$$

[7] This comes from the fact that for any Laurent polynomial $P(z)$, both $P(z)P(z^{-1})$ and $P(z) + P(z^{-1})$ will be Laurent polynomials. Furthermore if $P_1(z)$ and $P_2(z)$ are Laurent polynomials then $P_1(z)P_2(z)$ and $P_1(z) + P_2(z)$ will be Laurent polynomials as well.

$$L_q(z) = \prod_{j=1}^{n}(1 - z/b_j).$$

Using Lemma 3 and its proof, we can factorize $O(z)$ as

$$O(z) = R(z)R(1/z)C(z)C(1/z) \tag{98}$$

such that

$$C(z) = \prod_{j=1}^{c/2}(1 - z/r_j) \tag{99}$$

$$R(z) = \prod_{j=c+1}^{c+re/2}(1 - z/r_j) \tag{100}$$

Letting $v_p = p_n \prod_{j=1}^{m}(-a_j)$ and $v_q = q_n \prod_{j=1}^{n}(-b_j)$, the spectral density of $\{Z_t\}$ can be expressed as

$$f_Z(\lambda) = S_Z(e^{-i\lambda}) = \frac{v_p}{2\pi v_q}\left|\frac{L_p(e^{-i\lambda})}{L_q(e^{-i\lambda})}\right|^2\left|R(e^{-i\lambda})\right|^2\left|C(e^{-i\lambda})\right|^2 \tag{101}$$

By Kolmogorov's Formula, the variance of the shocks in the Wold representation ($\sigma^2_{\epsilon_z}$) is given by

$$\sigma^2_{\epsilon_z} = 2\pi exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} lnf_Z(\lambda)d\lambda\right\} \tag{102}$$

Since

$$\int_{-\pi}^{\pi} lnf_Z(\lambda) = \int_{-\pi}^{\pi}\ln\left(\frac{v_p}{2\pi v_q}\right)d\lambda + \int_{-\pi}^{\pi}\ln\left|\frac{L_p(e^{-i\lambda})}{L_q(e^{-i\lambda})}\right|^2 d\lambda + \int_{-\pi}^{\pi}\ln\left|R(e^{-i\lambda})\right|^2\left|C(e^{-i\lambda})\right|^2 d\lambda \tag{103}$$

we will handle each of the three additive terms on the right-hand side separately. First note that

$$\int_{-\pi}^{\pi}\ln\left(\frac{v_p}{2\pi v_q}\right)d\lambda = 2\pi\ln\frac{v_p}{2\pi v_q} \tag{104}$$

Next we note that

$$\int_{-\pi}^{\pi} \ln\left|\frac{L_p(e^{-i\lambda})}{L_q(e^{-i\lambda})}\right|^2 d\lambda = \int_{-\pi}^{\pi}\Big[\sum_{j=1}^{m}\ln[(1-e^{-i\lambda}/a_j)(1-e^{i\lambda}/\bar{a}_j)] - \sum_{j=1}^{n}\ln[(1-e^{-i\lambda}/b_j)(1-e^{i\lambda}/\bar{b}_j)]\Big]d\lambda$$
$$= \sum_{j=1}^{m}\int_{-\pi}^{\pi}\ln[(1-e^{-i\lambda}/a_j)(1-e^{i\lambda}/\bar{a}_j)]d\lambda - \sum_{j=1}^{n}\int_{-\pi}^{\pi}\ln[(1-e^{-i\lambda}/b_j)(1-e^{i\lambda}/\bar{b}_j)]d\lambda$$
$$= -\sum_{j=1}^{m}\int_{-\pi}^{\pi}\Big(\sum_{k=1}^{\infty}\frac{\frac{1}{a_j^k}e^{-ik\lambda}}{k} + \sum_{k=1}^{\infty}\frac{\frac{1}{\bar{a}_j^k}e^{ik\lambda}}{k}\Big)d\lambda + \sum_{j=1}^{n}\int_{-\pi}^{\pi}\Big(\sum_{k=1}^{\infty}\frac{\frac{1}{b_j^k}e^{-ik\lambda}}{k} + \sum_{k=1}^{\infty}\frac{\frac{1}{\bar{b}_j^k}e^{ik\lambda}}{k}\Big)d\lambda$$
$$= 0 \qquad (105)$$

where we note that the Taylor-series expansions hold since $\left|\frac{1}{a_j}\right| < 1$ and $\left|\frac{1}{b_j}\right| < 1$ for all j.

Since any root r_j on the unit circle can be expressed as $e^{i\lambda_j}$, such that $\lambda_j \in [-\pi, \pi]$ and $C(e^{-i\lambda}) = \prod_{j=1}^{c/2}(1 - e^{-i(\lambda+\lambda_j})$ we observe that

$$\left|C(e^{-i\lambda})\right|^2 = \left|\prod_{j=1}^{c/2}\Big[1-(cos(\lambda+\lambda_j) - isin(\lambda+\lambda_j)\Big]\right|^2$$
$$= \prod_{j=1}^{c/2}\Big[[1-(cos(\lambda+\lambda_j)]^2 + sin^2(\lambda+\lambda_j)\Big]$$
$$= \prod_{j=1}^{c/2}\Big[1-2cos(\lambda+\lambda_j) + cos^2(\lambda+\lambda_j) + sin^2(\lambda+\lambda_j)\Big]$$
$$= \prod_{j=1}^{c/2}\Big[2-2cos(\lambda+\lambda_j)\Big]$$
$$= \prod_{j=1}^{c/2} 4sin^2\Big(\frac{\lambda+\lambda_j}{2}\Big) \qquad (106)$$

Therefore

$$\int_{-\pi}^{\pi}\ln\left|R(e^{-i\lambda})\right|^2\left|C(e^{-i\lambda})\right|^2 d\lambda =$$

$$= \int_{-\pi}^{\pi}\Big[\sum_{j=1}^{c/2}\ln\Big[4sin^2\Big(\frac{\lambda+\lambda_j}{2}\Big)\Big] + \sum_{j=c+1}^{c+re/2} ln[(1-e^{-i\lambda}/a_j)(1-e^{-i\lambda}a_j)]\Big]d\lambda$$
$$= \sum_{j=1}^{c/2}\int_{-\pi}^{\pi}\ln\Big[4sin^2\Big(\frac{\lambda+\lambda_j}{2}\Big)\Big]d\lambda + \sum_{j=c+1}^{c+re/2}\int_{-\pi}^{\pi} ln[(1-e^{-i\lambda}/a_j)(1-e^{-i\lambda}a_j)]d\lambda$$
$$= 0 + 0 = 0 \qquad (107)$$

where we use the fact that $\int_{-\pi}^{\pi} \ln(1-e^{-i\lambda})d\lambda = 0$ and that for any λ_j, $\int_{-\pi}^{\pi} \ln\left[4sin^2\left(\frac{\lambda_j+\lambda}{2}\right)\right]d\lambda =$ 0. The latter equality is obtained from:

$$\begin{aligned}
\int_{-\pi}^{\pi} \ln\left[4sin^2\left(\frac{\lambda_j+\lambda}{2}\right)\right]d\lambda &= 2\pi\ln[4] + \int_{-\pi}^{\pi} \ln\left[sin^2\left(\frac{\lambda_j+\lambda}{2}\right)\right]d\lambda \\
&= 2\pi\ln[4] + \int_{-\pi-\lambda_j}^{\pi-\lambda_j} \ln\left[sin^2\left(\frac{\lambda_j+\lambda}{2}\right)\right]d\lambda \\
&= 2\pi\ln[4] + \int_{-\pi}^{\pi} \ln\left[sin^2\left(\frac{y}{2}\right)\right]dy \\
&= 2\pi\ln[4] + \int_{-\pi}^{0} \ln\left[sin^2\left(\frac{y}{2}\right)\right]dy + \int_{0}^{\pi} \ln\left[sin^2\left(\frac{y}{2}\right)\right]dy \\
&= 2\pi\ln[4] + 2\int_{0}^{\pi} \ln\left[sin^2\left(\frac{y}{2}\right)\right]dy \\
&= 2\pi\ln[4] + 4\int_{0}^{\pi} \ln\left[sin\left(\frac{y}{2}\right)\right]dy \\
&= 2\pi\ln[4] + 8\int_{0}^{\frac{\pi}{2}} \ln[sin(x)]dx \\
&= 2\pi\ln[4] - 8(\pi/2)\ln[2]) \\
&= 0 \qquad (108)
\end{aligned}$$

Thus using (103) we have

$$\int_{-\pi}^{\pi} lnf_Z(\lambda) = 2\pi\ln\frac{v_p}{2\pi v_q} \qquad (109)$$

and from (102) we obtain

$$\begin{aligned}
\sigma^2_{\epsilon_z} &= 2\pi exp\{\frac{1}{2\pi}\cdot 2\pi\ln\frac{v_p}{2\pi v_q}\} && (110) \\
&= 2\pi exp\{\ln\frac{v_p}{2\pi v_q}\} && (111) \\
&= 2\pi\frac{v_p}{2\pi v_q} && (112) \\
&= \frac{v_p}{v_q} && (113) \\
&= \frac{p_m\prod_{j=1}^{m}(-a_j)}{q_n\prod_{j=1}^{n}(-b_j)}. && (114)
\end{aligned}$$

□