Updating Forecasts in Vector Autoregression Models with an Application to the Canadian Banking Industry

Julia Hui Zhu *
with Allan W. Gregory

Department of Economics, Queen's University†

Job Market Paper

March 14, 2009

Abstract

Forecasting firms' earnings has long been an interest of market participants and academics. Traditional forecasting studies in a multivariate time series setting do not take into account that the timing of data release for a specific time period of observation is often spread over several days or weeks. This paper focuses on the separation of announcement timing or data release and the use of econometric real-time methods, which we refer to as an updated vector autoregression (VAR) forecast, to predict data that have yet to be released. In comparison to standard time series forecasting, we show that the updated forecasts will be more accurate the higher the correlation coefficients among the standard VAR innovations are. Forecasting with the sequential release of information has not been studied in the VAR framework, and our approach to the six Canadian banks shows its value. By using the updated VAR forecast, we find that the relative efficiency gain is 33% in the one-step-ahead forecast compared to the ordinary VAR forecast, and 7% compared to professional consensus forecasts. Thought experiments suggest that if banks' order of information release were to change, forecast errors could be substantially reduced. These experiments emphasize that evaluating the release ordering is crucial in determining forecast accuracy.

---

1 Introduction

The timing of data release for a specific time period of observation is often spread over weeks. For instance, earning announcements for firms can be spread over a two-week period, even though these earnings are for the same month or quarter. Further, official government statistics are often released at different times over the month or quarter and yet cover the same time period. This paper focuses on this separation of announcement timing or data release and the use of standard econometric updating methods to forecast data that are not yet available. To the best of our knowledge, this important aspect of forecasting has a direct application in financial market updating (in terms of earnings, earning per share, and so on).

Traditional forecasting in a multivariate time series setting is usually studied in the context of vector autoregression (VAR) models. In this set-up the VAR, with a common end-point, is specified and estimated. Single or multiple period forecasts can then be made. Standard errors for these forecasts can be based on asymptotic normal theory or, more recently, can be obtained using bootstrap or some other re-sampling techniques. This type of forecasting assumes that there is a sequence of common end-points of observed data and that forecasts are made for all variables over the same forecasting horizon. We consider a situation different from the standard VAR and data collection set-up in that only some of the variables comprising the system are released at a given point in time, with the remaining variables being released at later dates. These later dates may coincide for some variables, or they may differ, in which case there is a sequence of release times for the system as a whole. When all variable observations are released we are in the usual VAR forecasting with a common end-point and no updating will occur until the next time period with sequential information release.

To facilitate the analysis, we make two critical exogeneity assumptions. The first assumption is that the timing order of the release of information, either earlier or later in the release cycle, does not depend on the information that is released. That is, if firms have poor earnings, we assume this does not influence the announcement date. From what we are able to tell about earnings announcements, the decision as to when to announce is made long in advance of the time when the earnings information would be credibly known to the firm, so this is unlikely to be an issue. The second exogeneity assumption is that the announcement of one firm's earnings on a given day will not affect the announcement of the earnings of a related firm on a later day. That is, if one firm announces large earnings in, say, the first quarter of 2008 on February 27, 2008, a related firm does not change its announced earnings for the same quarter when released on, say, February 28, 2008. In the present application to the big six Canadian banks, one would expect that a great deal of time is required to

calculate earnings so that firms are unlikely to engage in such strategic interaction over such a short time horizon.

After Sims's (1980) influential work, VARs are widely applied in analyzing the dynamics of economic systems. For over twenty years, multivariate VAR models have been proven to be powerful and reliable tools. Stock and Watson (2001) reassess how well VARs have addressed data description, forecasting, structural inference, and policy analysis. Working with the inflation-unemployment-interest rate VAR, they conclude that VAR models either do no worse than or improve upon univariate autoregressive models and that both improve upon the random walk forecast. Therefore, VAR models are now rightly used in data description and forecasting. Generally, the ordinary multivariate VAR forecast is based on a common end-point. Hence, the standard VAR forecast does not distinguish that the timing of data release is often spread over weeks for the same time period of observation. However, most available macroeconomic variables or financial time series for a given time period are released on different announcement dates. For example, financial time series such as firms' earnings for the same quarter or the same year are made available at different dates for public use; macroeconomic indicators, such as inflation, employment, unemployment rate, interest rate, and so on, released for the same month, are also available to the public at different dates. Therefore, one objective of this research is to employ the familiar Kalman filtering set-up with prediction and updating. This discussion is specialized to the VAR framework, which we believe is new.

Not surprisingly, in comparison to the VAR forecast with a common end-point, the use of new sequential information leads to lowering of the mean squared error (MSE). We first consider a case with one available real-time variable in the updated bivariate VAR forecasts. When one variable's information is released, the difference from the release value and the prediction from the VAR represents the "new" information. If this new information is correlated with the innovation of the remaining variables, then there can be a meaningful update. When the correlation coefficient of the innovations between two variables is lower, the updated bivariate VAR forecast has a relatively smaller MSE than the ordinary bivariate VAR forecast. When the correlation coefficient of innovations between two variables is high, the updated bivariate VAR forecast has a smaller MSE than the ordinary bivariate VAR forecast. In the extreme case when the correlation coefficient approaches zero, then there is no improvement in the forecast. Of course, for multistep-ahead forecasts, having the additional information will always lead to improved forecasts with lower MSE.

The second scenario we consider is to update the bivariate VAR forecast with two more periods of real-time information known in advance. We find that, given a certain sufficiency condition, it is more accurate to adopt the new information in the VAR forecast. Inter-

estingly, there are instances when forecasts using the new information can actually lead to poorer predicted outcomes.

Finally, we conduct the extension to the multivariate VAR forecast. With one available real-time variable known in advance, the methodology used to obtain the MSE of the remaining variables is straightforward and follows the same direction as the bivariate case. With two more available real-time variables known in advance, we use the innovations of all available variables to predict the innovation of each of the remaining variables. We show that the forecast MSE for a variable monotonically decreases as other variables that are related are sequentially released.

To illustrate these methods, we choose earnings per share (EPS)[1] of the big six Canadian banks. These six banks dominate the market with 90% of all banking assets under their control. The banks are: Royal Bank of Canada (RBC), Bank of Nova Scotia (BNS), Toronto-Dominion Bank (TD), Bank of Montreal (BMO), Canadian Imperial Bank of Commerce (CIBC), and National Bank of Canada (BNC). These banks' quarterly EPS figures and the corresponding analyst forecast data are well documented and accessible. We collect the data from the Institutional Brokers Estimate System (I/B/E/S) database provided by Wharton Research Data Service and from the current quarterly financial statements of each bank. We have obtained the past 23 years of quarterly banks' earnings per share and individual analysts' forecast of EPS data dating from the second quarter of 1986 to the first quarter of 2009. We compare to the consensus earnings forecast which is available for all these banks and is calculated as the average of all available individual analysts' forecasts. The big six Canadian banks are all cross-listed between the New York Stock Exchange and the Toronto Stock Exchange. Under the U.S. Securities and Exchange Commission, the filing deadline for quarterly reports is 35 days, whereas under the Ontario Securities Commission, the deadline for filing interim financial statements is 45 days. Overall, for the big six Canadian banks, quarterly financial announcements are mandatory within a certain time range; the timing

---

[1]http://www.martinmarietta.com/Investors/Non-GAAP.asp

The U.S. Securities and Exchange Commission defines a "non-GAAP financial measure" as a numerical measure of historical or future financial performance, financial positions, or cash flows that excludes amounts, or is subject to adjustments that effectively exclude amounts, included in the most directly comparable measure calculated and presented in accordance with GAAP in financial statements, and vice versa for measures that include amounts, or is subject to adjustments that effectively include amounts, that are excluded from the most directly comparable measure so calculated and presented. For these purposes, "GAAP" refers to generally accepted accounting principles in the United States. Non-GAAP financial measures disclosed by management are provided as additional information to investors in order to provide them with an alternative method for assessing firms' financial condition and operating results. These measures are not in accordance with, or a substitute for, GAAP, and may be different from or inconsistent with non-GAAP financial measures used by other companies. The non-GAAP measures are used in banks' own management discussion and analysis and, therefor, may not be comparable to similar terms used by other banks.
To be consistent with banks' own management and analysis, we use non-GAAP diluted EPS in this paper.

3

of announcements varies, and the order of announcements is different at each quarter. We find that the root mean squared error in the updated VAR is less than the ordinary VAR. This shows quantitatively that the updated VAR improves 33% in a one-step-ahead forecast compared to the ordinary VAR, and improves 7% in comparison to consensus forecasts.

Obviously, market participants follow firms' current and expected future earnings. Brown and Rozeff (1978) and Bown, Hagerman, and Zmijewski (1987) conclude that analysts' forecasts significantly outpredict time series model forecasts, while Cragg and Malkiel (1968), Elton and Gruber (1972), and Imhoff and Pare (1982) contend that analysts' earnings forecasts are not significantly more accurate than time series model forecasts. Hopwood, McKeown and Newbold (1982) find that time series methods outperform a random walk by using quarterly earnings information to predict annual EPS. These conflicting findings have led Conroy and Harris (1987) to conclude that a combination of analysts' and time series annual EPS forecasts may provide superior earnings forecasts, especially with a few analysts' forecasts. Individual time series methods of predicting EPS based on past annual EPS have done no better than simple random walk models. Recently, substantial attention has been paid to the characteristics of analysts' forecasts of earnings. More specifically, researchers document that analysts' forecasts of earnings are biased (Beckers, Steliaros and Thomas 2004, Hong and Kubik 2003), inefficient (Easterwood and Nutt 1999), and irrational (De Bondt and Thaler 1990, Aborbanell and Bernard 1992, and Lim 2001), and indicate herding behavior (Clement and TSE 2005). However, Keane and Runkle (1998) use methods of moments tests and find that analysts' forecasts of corporate profits are rational, which challenges the earlier work of De Bondt and Thaler (1990) and Aborbanell and Bernard (1992).

In this study, the order of release of information is exogenous by assumption. We can, however, construct any alternative orderings we wish and rank ordering according to some criterion. A natural criterion is an ordering that maximize disclosure. The objective function is to minimize mean squared forecast errors to provide the market with the most informative release. Our theoretical results and simulation findings indicate that the lowest mean squared forecast error does exist, depending on firms' variance and their correlation coefficients. As a result, we investigate how the timing of earnings announcements is helpful in accurately forecasting firms' earnings. Thought experiments explore the order of earnings release for the big six Canadian banks to maximize disclosure. Under this ordering, a substantial decrease in MSE can be obtained in comparison to the actual ordering.

The remainder of the paper is organized as follows. Section 2 presents our econometric methodology. Section 3 reports the simulation studies. An application of the updated VAR forecasts to the Canadian banking industry and thought experiments are illustrated in section 4. Section 5 summarizes our main findings. The appendix provides the proofs.

2 A Theoretical Framework

Consider the N multivariate stationary VAR (1) model

$$Y_t = AY_{t-1} + \epsilon_t, \qquad t = 0, \pm 1, \pm 2, ..., \qquad (1)$$

where $Y_t = (y_{1t}, ..., y_{Nt})'$ is a $(N \times 1)$ random vector, and the A is fixed $(N \times N)$ coefficient matrices. The first subscript represents the variable and the second subscript represents the time period. Moreover, $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{Nt})'$ is a N-dimensional white noise or innovation process, that is, $E(\epsilon_t) = 0$ and $E(\epsilon_t \epsilon_t') = \Omega_\epsilon$, with the contemporaneous covariance $Cov(\epsilon_{it}, \; \epsilon_{jt}) = \rho_{ij} \sigma_i \sigma_j$ for $i = 1, ..., N$ and $j = 1, ..., N$. The covariance matrix Ω_ϵ is assumed to be nonsingular if not otherwise stated. Since we assume stationary vector autoregressive process, the condition of correlation coefficient $|\rho_{ij}| < 1$ must hold. Finally, σ_i is the standard deviation of the innovation ϵ_i. Given the multivariate VAR model (1), the ordinary multivariate VAR one-step-ahead forecast at time region T is $Y_T(1) = A \, Y_T$ and the associated forecast error is $\epsilon_T(1) = \epsilon_{T+1}$, where $Y_T(1)$ denotes the forecast of Y at time $T+1$ and $\epsilon_T(1)$ denotes forecast error at time $T+1$. The ordinary multivariate VAR forecasts are standard and can be obtained from Lütkepohl (1993) and Hamilton (1994).

Theoretically, the multivariate VAR forecast is based on the certain amount of time periods of 1 through T. Each equation in the multivariate VAR model can be estimated by ordinary least squares (OLS) regression. This OLS estimator is as efficient as the maximum likelihood estimator and the general least squared estimator. Therefore, the ordinary multivariate VAR forecast computed through the unbiased and consistent coefficient estimates and the variance covariance matrix estimates has the lowest MSE and is optimal. However, the fact is that most macroeconomic or financial time series we study in multivariate VAR forecast do not end at the same time, that is, one variable is generally available for public use a couple of days prior to the other variables. For example, financial time series such as firms' earnings for the same quarter or the same year are made available at different dates for public use; macroeconomic indicators such as inflation, employment, unemployment rate, interest rate, and so on, released for the same month are also available at different dates for public use. Omitting the timing factor, the ordinary multivariate VAR forecast usually ignores the latest information we can obtain and adopts the same amount of certain time periods to make forecasts.

The focus of this paper is to examine how taking advantage of more data from one variable and matching the timing factor can be used to improve multivariate VAR forecasts.

2.1 Updating Bivariate VAR Forecast

A practical method to update forecasts with one real-time variable available in advance is investigated. To simplify the discussion, we first consider the bivariate VAR forecast, where $N = 2$, firstly. Suppose y_{1t} is observable for $t = 1, \ldots T + 1$ and y_{2t} is observable only up to time T. Then the reduced form bivariate VAR (1) model is as follows

$$
\begin{aligned}
y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + \epsilon_{1t} && t = 1, ..., T, T+1 && (2) \\
y_{2t} &= a_{21}y_{1t-1} + a_{22}y_{2t-1} + \epsilon_{2t} && t = 1, ..., T.
\end{aligned}
$$

Following the ordinary multivariate VAR forecasting method proposed by Lütkepohl (1993), the one-step-ahead ordinary bivariate VAR forecast error covariance matrix (or forecast MSE matrix) is

$$
MSE[Y_T(1)] = \Omega_\epsilon,
$$

where Y is a vector of $(y_1, y_2)'$, and the covariance matrix of Ω_ϵ is $E(\epsilon_t, \epsilon_t')$; that is,

$$
\Omega_\epsilon = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.
$$

Therefore, the MSE of one-step-ahead forecast of variable y_2 in the ordinary bivariate VAR is $MSE[y_{2T}(1)] = \sigma_2^2$. Hereafter we denote MSE^u as the updated multivariate VAR forecast MSE to distinguish it from the ordinary multivariate VAR forecast MSE.

With one more piece of real-time information (y_{1T+1}) available in advance, at the forecast horizon of the time period $T + 1$, we observe ϵ_{1T+1}. This is due to $\epsilon_{1T+1} = y_{1T+1} - (a_{11}y_{1T} + a_{12}y_{2T})$. If we regress ϵ_2 on ϵ_1, then the best estimates of ϵ_{2T+1} is obtained by $\epsilon_{2T+1|T} = E(\epsilon_{2T+1}|\epsilon_{1T+1})$. Thus, we can forecast the residual $\epsilon_{2T+1|T}$ by the relationship $\epsilon_{2T+1|T} = (\rho_{12}\sigma_2/\sigma_1)\epsilon_{1T+1}$. Hence, the one-step-ahead forecast MSE is obtained as the following proposition:

Proposition 1 *Given the full information set $\mathcal{I}_1 = \{y_{11}, y_{12}, ..., y_{1T+1}, y_{21}, ..., y_{2T}\}$, the mean squared error of one-step-ahead forecast of variable y_2 in the updated bivariate VAR is*

$$
MSE^u[y_{2T}(1)] = (1 - \rho_{12}^2)\sigma_2^2 \tag{3}
$$

Proof. *See Appendix.*

This proposition shows that taking advantage of one more piece of real-time information available in advance implies that the one-step-ahead updated bivariate VAR forecast MSE^u

is $(1 - \rho_{12}^2)\sigma_2^2$. Since ρ_{12} is correlation between ϵ_1 and ϵ_2, the condition $|\rho_{12}| < 1$ must hold by stationary. However, omitting the extra information y_{1T+1}, the ordinary bivariate VAR forecast MSE gives $MSE = \sigma_2^2$. Therefore, the one-step-ahead updated bivariate VAR forecast has smaller MSE in comparison with the forecast from the ordinary bivariate VAR. This implies that the updated bivariate VAR forecast is more accurate than the forecast from the ordinary bivariate VAR.

Additionally, the higher the correlation among the error terms of the variables, the smaller the MSE in the updated bivariate VAR forecast. When the correlation coefficient of the innovations between two variables is lower, the updated bivariate VAR forecast has a relatively smaller MSE than the ordinary bivariate VAR forecast. When the correlation coefficient of observation innovations between two variables is higher, the updated bivariate VAR forecast has a significantly smaller MSE than the ordinary bivariate VAR forecast. In one extreme case, when the correlation coefficient approaches zero, that is, no useful contemporaneous information is available for forecasting, the updated bivariate VAR forecast has exactly the same results as the ordinary bivariate VAR forecast. In another extreme case, when the correlation coefficient approaches one, that is, we have perfect linear association and there are no errors, the updated bivariate VAR forecast has the best performance.

To generalize, we also examine the $k \geq 2$ long-horizon forecast with one more piece of real-time information known in advance.

Proposition 2 *Given the full information set $\mathcal{I}_1$, the k-step-ahead forecast mean squared error matrix in the updated bivariate VAR is*

$$
\begin{aligned}
MSE^u[Y_T(k)] &= \sum_{i=0}^{k-2} A^i \Omega_\epsilon A^{i'} + A^{k-1} \begin{pmatrix} 0 & 0 \\ 0 & (1 - \rho_{12}^2)\sigma_2^2 \end{pmatrix} A^{k-1'} \qquad (4) \\
&= \Omega_\epsilon + A \ \ MSE^u[Y_T(k-1)] \ \ A', \qquad k \geq 2,
\end{aligned}
$$

where A is 2×2 dimensional coefficient matrix. A matrix to the power of zero is defined to be the identity matrix of the same dimensions, that is, $A^0 = I$, and

$$
A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.
$$

Proof. *See Appendix.*

This proposition shows that the multistep-ahead updated bivariate VAR forecast builds upon the first step forecast derived from proposition (1). By iterating forward, we see that the k-step-ahead updated bivariate VAR forecast mean squared error matrix MSE^u in equation

(4) is smaller than the recursive ordinary bivariate VAR forecast MSE matrix

$$MSE[Y_T(k)] = \sum_{i=0}^{k-2} A^i \Omega_\epsilon (A^i)' + A^{k-1} \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} A^{k-1'} \tag{5}$$
$$= MSE[Y_T(k-1)] + A^{k-1}\Omega_\epsilon A^{k-1'} \quad k \geq 2.$$

To see the difference between the updated bivariate VAR forecast MSE^u and the ordinary bivariate VAR forecast MSE, we compare the equation (4) with (5):

$$MSE[Y_T(k)] - MSE^u[Y_T(k)] = A^{k-1} \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \rho_{12}^2\sigma_2^2 \end{pmatrix} A^{k-1'}. \tag{6}$$

In a one-step-ahead forecast ($k = 1$) equation (6) indicates that the gain of employing one available real-time information y_{1T+1} to forecast related variable y_{2T+1} depends on the correlation coefficient of the innovations, ρ_{12}. When $\rho_{12} = 0$, there is no gain at all; the updated bivariate VAR forecast has the same MSE as the ordinary bivariate VAR forecast. This implies that the updated bivariate VAR forecast is more accurate than the forecast from the ordinary bivariate VAR under short-horizon forecasting. Furthermore, the higher the correlation among the error terms of the variables, the smaller the MSE in the updated bivariate VAR forecast. When the correlation coefficient approaches zero, that is, no contemporaneous information is available to be useful for forecasting, the updated forecast has exactly the same results as the ordinary bivariate VAR forecast. When the correlation coefficient approaches one, we have perfect linear association; there are no errors and the updated bivariate VAR forecast has the best performance. As forecasting horizon becomes longer, the MSE of the updated bivariate VAR forecast converges to the MSE of the ordinary bivariate VAR forecast.

In a two-step-ahead forecast ($k = 2$) equation (6) becomes

$$MSE[Y_T(2)] - MSE^u[Y_T(2)] = A \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \rho_{12}^2\sigma_2^2 \end{pmatrix} A'$$
$$= \begin{pmatrix} (a_{11}\sigma_1 + a_{12}\rho_{12}\sigma_2)^2 & (a_{11}\sigma_1 + a_{12}\rho_{12}\sigma_2)(a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2) \\ (a_{11}\sigma_1 + a_{12}\rho_{12}\sigma_2)(a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2) & (a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2)^2 \end{pmatrix}.$$

Since the difference between the forecast MSE of y_{2T+1} of the ordinary bivariate VAR and that of the updated bivariate VAR is $(a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2)^2$, this is larger or equal to zero. Therefore, the two-step-ahead updated bivariate VAR forecast has smaller MSE compared with the forecast from the ordinary bivariate VAR. Even with $\rho_{12} = 0$, the gain of employing

one available real-time information y_{1T+1} to forecast the related variable y_{2T+1} still exists since $a_{21}^2 \sigma_1^2$ is always positive.

Furthermore, $MSE[Y_T(k)] - MSE^u[Y_T(k)]$ converges to zero rapidly as $k \to \infty$. In other words, the MSE by the updated VAR forecast converges to the MSE by the ordinary VAR forecast as the forecasting horizon gets larger. Under the assumption that our VAR (1) process is stationary, the polynomial $det(I_m - Az)^2$ has no roots in and on the complex unit circle. That is equivalent to say that all eigenvalues of parameter matrix A have modulus less than 1. By the properties of matrices[3], A^{k-1} converges to 0 as $k \to \infty$.

2.2 Updating Bivariate VAR Forecast with Data from Two More Periods Known in Advance

In practice, there are a number of applications of the bivariate VAR with two more periods of available real-time information known in advance. For instance, in Canada, Telehealth is a toll-free helpline provided by the Ontario Ministry of Health and Long-term Care's Telehealth program and is available to all residents of Ontario. Users are encouraged to call with any general health questions with confidential advice being given regarding any health concerns. National Ambulatory Care Reporting System (NACRS) was developed in 1997 by the Canadian Institute for Health Information to capture clinical, administrative and demographic information from all hospital-based and community-based ambulatory care. The concern with these two correlated sources of data is one of timeliness, as NACRS data is not available in real-time, but rather months later. This issue is also compounded by the fact that some hospitals have yet to complete a migration to electronic records management, making the integration of all NACRS data additionally difficult. The goal is to use the Telehealth data as a forecast to alert physicians under flu outbreak or vaccine failure has occurred and high demand for NACRS data is expected. Understanding the impact of these limitations is crucial to study the provincial Telehealth data and its usefulness to public health and emergency services.

In this section, we develop the updated bivariate VAR forecast with two more time periods of real-time information ($s \geq 2$). Let y_{1t} be observable for $t = 1, \ldots T + s$ with $s \geq 2$ and let y_{2t} be observable only up to time T. The simple bivariate VAR(1) model is as

---

[2]Lütkepohl (1991) Appendix A.6 rule 7 on page 456: all eigenvalues of the $(m \times m)$ matrix A have modulus less than 1 if and only if $det(I_m - Az) \neq 0$ for $|z| \leq 1$, that is, the polynomial $det(I_m - Az)$ has no roots in and on the complex unit circle.

[3]Lütkepohl (1996) property 14 on page 39: A $(m \times m)$ matrix A, $A^i \to_{i \to \infty} 0 \Leftrightarrow$ all eigenvalues of A have modulus less than 1.

follows:

$$y_{1t} = a_{11}y_{1t-1} + a_{12}y_{2t-1} + \epsilon_{1t} \qquad t = 1, ..., T, T+1, ..., T+s \qquad (7)$$
$$y_{2t} = a_{21}y_{1t-1} + a_{22}y_{2t-1} + \epsilon_{2t} \qquad t = 1, ..., T.$$

We assume all the assumptions hold as in the section 2.1.

A one-step-ahead forecast MSE in this model set up is consistent with the proposition 3. The MSE^u in this updated bivariate VAR forecast given one more period of available real-time information is smaller than the ordinary bivariate VAR forecast, that is, $MSE^u = (1 - \rho_{12}^2)\sigma_2^2 < MSE = \sigma_2^2$.

For a two-step-ahead forecast, rearranging equation (7), we obtain

$$\epsilon_{1T+2} = y_{1T+2} - (a_{11}y_{1T+1} + a_{12}y_{2T+1}).$$

Since we do not observe y_{2T+1}, we do not observe ϵ_{1T+2}. There are two ways to make a prediction on ϵ_{1T+2}. One way is simply to set the unconditional prediction $E(\epsilon_{1T+2}) = 0$ with variance ϵ_{1T+2} (σ_1^2) the first element of the MSE of the ordinary bivariate VAR forecast, that is, the first row and column element of the matrix $(\Omega_\epsilon + A\Omega_\epsilon A')$. The alternative way is to predict $\epsilon_{1T+2|T}$ through the residual form $\epsilon_{1T+2|T} = y_{1T+2} - (a_{11}y_{1T+1} + a_{12}y_{2T+1|T})$. The question is wether the conditional predictor of ϵ_{1T+2} has a larger MSE than the unconditional predictor with MSE of σ_1^2. In the former case, the variance of the difference in error becomes

$$Var[\epsilon_{1T+2} - \epsilon_{1T+2|T}] = a_{12}^2(1 - \rho_{12}^2)\sigma_2^2.$$

If a sufficient condition $\{a_{12}^2(1 - \rho_{12}^2)\sigma_2^2 <$ the first row and column element of matrix $\Omega_\epsilon + A\Omega_\epsilon A'\}$ holds, we would use $\epsilon_{1T+2|T}$ rather than $E(\epsilon_{1T+2}) = 0$. Then the variance of the forecast error is

$$Var[y_{2T+2} - y_{2T+2|T}] = (1 + a_{22}^2)(1 - \rho_{12}^2)\sigma_2^2.$$

We see that not only the correlation coefficient of two time series and the variance of the forecasting innovation, but also the coefficient parameter a_{22} plays a role in the multistep-ahead forecast.

To generalize the standard $2 \leq k \leq s$ multistep horizon forecast, we iterate to obtain the k-step-ahead updated bivariate VAR forecast.

Proposition 3 *Given the full information set $\mathcal{I}_2 = \{y_{11}, y_{12}, ..., y_{1T+1}, ..., y_{1T+s}, y_{21}, ..., y_{2T}\}$, and the sufficient condition $\{(\sum_{i=0}^{k-2} a_{22}^{2i})a_{12}^2\sigma_2^2(1 - \rho_{12}^2) <$ the first row and column element of matrix $\Sigma_{i=0}^{k-1} A^i \Omega_\epsilon A^{i'}\}$ holds. Then the k-step-ahead forecast mean squared error matrix in*

the updated bivariate VAR is

$$MSE^u[y_{2T}(k)] = (\sum_{i=0}^{k-1} a_{22}^{2i})(1 - \rho_{12}^2)\sigma_2^2 \qquad 2 \le k \le s. \tag{8}$$

Proof. *See Appendix.*

Clearly, if the absolute value of estimated a_{22} is larger than one, the MSE of the updated bivariate VAR forecast diverges when the forecast horizon is large enough, that is as $k \to \infty$. However, the stationary assumption rules out the divergence case. Therefore, when the forecast horizon is large enough, the MSE of the updated bivariate VAR forecast will converge to the MSE of the ordinary bivariate VAR forecast. If the absolute value of estimated a_{22} is less or equal to one, there exists a permanent efficient gain by employing the updated bivariate VAR forecast.

2.3 Updating Multivariate VAR Forecast

In the last two sections, we analyze the different kinds of the updated bivariate VAR forecasts. We now extend our investigation to the multivariate case, where a separate sequence of informational release occurs.

One Available Real-time Variable

Let y_{1t} be a real-time variable observed for $t = 1, \ldots, T + 1$ and $y_{2t}, y_{3t}, \ldots, y_{Nt}$ be observable only up to time T. Then a multivariate VAR model (1) is

$$
\begin{aligned}
y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + \ldots + a_{1N}y_{Nt-1} + \epsilon_{1t} & t &= 1, ..., T+1 & (9)\\
y_{2t} &= a_{21}y_{1t-1} + a_{22}y_{2t-1} + \ldots + a_{2N}y_{Nt-1} + \epsilon_{2t} & t &= 1, ..., T \\
&\quad \ldots & t &= 1, ..., T \\
y_{Nt} &= a_{N1}y_{1t-1} + a_{N2}y_{2t-1} + \ldots + a_{NN}y_{Nt-1} + \epsilon_{Nt} & t &= 1, ..., T.
\end{aligned}
$$

The updating procedure is as before. Given the known full information set $\mathcal{I}_3 = \{y_{11}, y_{12}, ..., y_{1T+1}, y_{21}, ..., y_{2T}, ..., y_{N1}, ...y_{NT}\}$, the mean squared error of one-step-ahead forecast of variable y_2 in the updated multivariate VAR is

$$MSE^u[y_{nT}(1)] = (1 - \rho_{1n}^2)\sigma_n^2 \qquad n \ge 2. \tag{10}$$

m More Available Real-time Variables

<div align="center">11</div>

We now move on to derive the updated multivariate VAR forecast with m more available real-time variables. First, we consider $m = 2$ case, that is two variables can be observed currently and we forecast the rest of $N - m$ variables at the same time period.

Let y_{1t} and y_{2t} be two real-time variables observed for $t = 1, \ldots, T+1$ and $y_{3t}, y_{4t}, \ldots, y_{Nt}$ be observable only up to time T. Then the multivariate VAR model (1) can be expanded as the following reduced form VAR (1) process

$$
\begin{aligned}
y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + \ldots + a_{1N}y_{Nt-1} + \epsilon_{1t} & t &= 1, \ldots, T+1 & (11)\\
y_{2t} &= a_{21}y_{1t-1} + a_{22}y_{2t-1} + \ldots + a_{2N}y_{Nt-1} + \epsilon_{2t} & t &= 1, \ldots, T+1 & (12)\\
y_{3t} &= a_{31}y_{1t-1} + a_{32}y_{2t-1} + \ldots + a_{3N}y_{Nt-1} + \epsilon_{3t} & t &= 1, \ldots, T\\
&\quad \ldots & t &= 1, \ldots, T\\
y_{Nt} &= a_{N1}y_{1t-1} + a_{N2}y_{2t-1} + \ldots + a_{NN}y_{Nt-1} + \epsilon_{Nt} & t &= 1, \ldots, T.
\end{aligned}
$$

We use the known information $\mathcal{I}_4 = \{y_{11}, \ldots, y_{1T}, y_{1T+1}; y_{21}, \ldots, y_{2T}, y_{2T+1}; y_{31}, \ldots, y_{3T}; \ldots; y_{N1}, \ldots, y_{NT}\}$ to forecast $y_{3T}(1), \ldots, y_{NT}(1)$. Equations (11) and (12) give

$$
\epsilon_{1T+1} = y_{1T+1} - a_{11}y_{1T} - a_{12}y_{2T} - \ldots - a_{1N}y_{NT}
$$
$$
\epsilon_{2T+1} = y_{2T+1} - a_{21}y_{1T} - a_{22}y_{2T} - \ldots - a_{2N}y_{NT}.
$$

As before, since we observe y_{1T+1}, y_{2T+1}, y_{1T}, y_{2T}, ..., y_{NT}, we know ϵ_{1T+1} and ϵ_{2T+1}. Similarly, assume all parameters are known in the coefficient matrices A and the covariance $Cov(\epsilon_{it}, \epsilon_{jt})$. The prediction equation is for updating the forecast. The third equation error can be calculated using the following regression.

$$
\epsilon_{3t} = \alpha_1 \, \epsilon_{1t} + \alpha_2 \, \epsilon_{2t} + \eta_t \qquad t = 1, \ldots T,
$$

where the α_1 and α_2 are the population least squares coefficients. We project the ϵ_{3t} onto both ϵ_{1t} and ϵ_{2t}. This yields least squares estimate of α:

$$
\begin{aligned}
\alpha_1 &= \frac{Var(\epsilon_2)Cov(\epsilon_1, \epsilon_3) - Cov(\epsilon_1, \epsilon_2)Cov(\epsilon_2, \epsilon_3)}{Var(\epsilon_1)Var(\epsilon_2) - Cov^2(\epsilon_1, \epsilon_2)}\\
&= (\frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{12}^2})\frac{\sigma_3}{\sigma_1}
\end{aligned}
$$

and

$$\begin{aligned} \alpha_2 &= \frac{Var(\epsilon_1)Cov(\epsilon_2,\epsilon_3) - Cov(\epsilon_1,\epsilon_2)Cov(\epsilon_1,\epsilon_3)}{Var(\epsilon_1)Var(\epsilon_2) - Cov^2(\epsilon_1,\epsilon_2)} \\ &= (\frac{\rho_{23} - \rho_{12}\rho_{13}}{1 - \rho_{12}^2})\frac{\sigma_3}{\sigma_2}. \end{aligned}$$

Hence, the best predictor of ϵ_{3T+1} is

$$\epsilon_{3T+1|T} = (\frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{12}^2})\frac{\sigma_3}{\sigma_1}\epsilon_{1T+1} + (\frac{\rho_{23} - \rho_{12}\rho_{13}}{1 - \rho_{12}^2})\frac{\sigma_3}{\sigma_2}\epsilon_{2T+1},$$

where $\rho_{13} = \rho_{31}$ and $\rho_{23} = \rho_{32}$ are the correlation coefficients of innovation ϵ_1 and ϵ_3, and ϵ_2 and ϵ_3 respectively. The forecast error for forecasting y_{3T+1} then becomes

$$y_{3T+1} - y_{3T+1|T} = \epsilon_{3T+1} - (\frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{12}^2})\frac{\sigma_3}{\sigma_1}\epsilon_{1T+1} - (\frac{\rho_{23} - \rho_{12}\rho_{13}}{1 - \rho_{12}^2})\frac{\sigma_3}{\sigma_2}\epsilon_{2T+1}.$$

To distinguish the notation used here with the notation used in the previous sections, we adopt the superscript to denote the order of the informational release. For instance, the corresponding mean squared error for forecasting y_{3T+1} using both variable 1 and 2 releases is $MSE^u[y_{3T}^{1:2}(1)]$. Moreover, this is equal to

$$MSE^u[y_{3T}^{1:2}(1)] = Var[y_{3T+1} - y_{3T+1|T}] = (1 - \frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{12}^2})\sigma_3^2, \qquad (13)$$

where $|\rho_{12}| \neq 1$.

To generalize the updated multivariate VAR forecast with m more available real-time variables, let y_{mt} be m real-time variables observed for $t = 1, \ldots, T+1$ and y_{nt} be observable only up to time T. Then the multivariate VAR model (1) can be expanded as follows

$$\begin{aligned} y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + \ldots + a_{1N}y_{Nt-1} + \epsilon_{1t} & t &= 1, \ldots, T+1 & (14) \\ &\cdots & t &= 1, \ldots, T+1 \\ y_{mt} &= a_{m1}y_{1t-1} + a_{m2}y_{2t-1} + \ldots + a_{mN}y_{Nt-1} + \epsilon_{mt} & t &= 1, \ldots, T+1 & (15) \\ y_{m+1,t} &= a_{m+1,1}y_{1t-1} + a_{m+1,2}y_{2t-1} + \ldots + a_{m+1,N}y_{Nt-1} + \epsilon_{m+1,t} & t &= 1, \ldots, T \\ &\cdots & t &= 1, \ldots, T \\ y_{Nt} &= a_{N1}y_{1t-1} + a_{N2}y_{2t-1} + \ldots + a_{NN}y_{Nt-1} + \epsilon_{Nt} & t &= 1, \ldots, T. \end{aligned}$$

Proposition 4 *Given the full information set $\mathcal{I}_5 = \{y_{11}, \ldots, y_{1T}, y_{1T+1}; \ldots; y_{m1}, \ldots, y_{mT}, y_{mT+1}; y_{m+1,1}, \ldots, y_{m+1,T}; \ldots; y_{N1}, \ldots, y_{NT}\}$, suppose the random vector $\epsilon_i = (\epsilon_1 \quad \epsilon_2 \quad \ldots \quad \epsilon_m)'$ and the*

random vector $\epsilon_j = (\epsilon_{m+1} \quad \epsilon_{m+2} \quad ... \quad \epsilon_N)'$ *are as follows*

$$\begin{pmatrix} \epsilon_i \\ \epsilon_j \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{\epsilon_i\epsilon_i} & \Omega_{\epsilon_i\epsilon_j} \\ \Omega_{\epsilon_i\epsilon_j} & \Omega_{\epsilon_j\epsilon_j} \end{pmatrix} \right).$$

Then the corresponding general form of mean squared error $MSE^u[y_{jT}^{1:2:....:m}(1)]$ *for forecasting* y_{jT+1} *is*

$$\begin{aligned} MSE^u[y_{jT}^{1:2:....:m}(1)] &= Var(y_{jT+1} - y_{jT+1|T}) \\ &= Var(\epsilon_{jT+1} - \Omega_{\epsilon_j\epsilon_i}\Omega_{\epsilon_i\epsilon_i}^{-1}\epsilon_{iT+1}) \\ &= Var(\epsilon_{jT+1}) + Var(\Omega_{\epsilon_j\epsilon_i}\Omega_{\epsilon_i\epsilon_i}^{-1}\epsilon_{iT+1}) \\ &\quad -2Cov(\epsilon_{jT+1}, \Omega_{\epsilon_j\epsilon_i}\Omega_{\epsilon_i\epsilon_i}^{-1}\epsilon_{iT+1}). \end{aligned} \tag{16}$$

where $\Omega_{\epsilon_i\epsilon_i}$, $\Omega_{\epsilon_j\epsilon_j}$, *and* $\Omega_{\epsilon_i\epsilon_j}$ *are the partition matrix. The matrix* $\Omega_{\epsilon_j\epsilon_i}$ *is the transpose matrix of* $\Omega_{\epsilon_i\epsilon_j}$.

Proof. *See Appendix.*

Given the information set $\mathcal{I}_3$, the forecasting MSE of variable y_{3T+1} by equation (10) is

$$MSE^u[y_{3T}^1(1)] = (1 - \rho_{13}^2)\sigma_3^2.$$

However, given the information set $\mathcal{I}_4$, the forecasting MSE of variable y_{3T+1} is as in (13). Comparing (10) with (13), we find that the forecasting MSE of variable y_{3T+1} is smaller when we use the most recent available information to forecast. That is, the outcome from equation (13) is smaller than the result of equation (10). To see how it happens, we can compare ρ_{13}^2 to $\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}/(1 - \rho_{12}^2)$.

$$\frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{12}^2} - \rho_{13}^2 = \frac{(\rho_{23} - \rho_{12}\rho_{13})^2}{1 - \rho_{12}^2} \geq 0.$$

Therefore, $MSE^u[y_{3T}^{1:2}(1)] < MSE^u[y_{3T}^1(1)]$. This comparison strengthens the usefulness of more pieces of information in forecasting.

Proposition 5 *Given the full information set* $\mathcal{I}_5$ *and the updated multivariate VAR model (14), the more pieces of available information we use, the smaller the forecast mean squared errors. In other words, for a given variable j*

$$MSE[y_{jt}^1(1)] \geq MSE[Y_{jt}^{1:2}(1)] \geq ... \geq MSE[Y_{jt}^{1:....:m-1}(1)] \geq MSE[Y_{jt}^{1:....:m}(1)] \tag{17}$$

14

Proof. *See Appendix.*

2.4   Sequential Arrival of Real-time Variables

A natural question arises in considering the sequential release of information: is there timing of announcement ordering that related firms follow in order to obtain the maximum disclosure. The answer to this question depends of course on what criterion one chooses to assess optimality. Suppose that each firm will release their information for the previous period separately; no two firms can release the same day (this may allow the market to focus on individual firms and information). An important economic earnings question worth addressing in an optimizing model is why firms do not all release earnings at the same time. What is the economic reason for separating and changing times for earnings releases.

Suppose we adopt an objective function that enforces a particular ordering of information with the rule that each firm must release data on a different order. We consider the question of order using the following criterion: information is released to minimize standardized mean squared forecast errors at each informational release (maximum total disclosure). The objective is to provide the market with the most informative release subject to not allowed to release information at the same time. Since the VAR structure is itself not related to this criterion, we can calculate for any given set of firms N, either of these, and compare them to actual practice.

In this set-up we have N separate informational releases and the problem can be solved. Given the nature of the interdependence in forecasting, the problem has to be solved for all combinations of information release dates over all time T. The order is assumed to be fixed over the T periods to make the problem meaningful. Otherwise, order would usually be chosen on the basis of what firm has the smallest forecast error at any time.

We minimize the criterion function for all T, so with N firms there are $N!$ (that is, $N * (N-1) * (N-2) * ... * 2 * 1$) different permutations or orderings. Assume that each firms earnings is normalized in some way (earnings per share or divide them by their standard deviation) and denote these as $(y_{1t}, y_{2t}, \ldots, y_{nt})$.

Consider the following criterion at release date T, which we minimize. Given the large number of potential orderings to search over, the notational set-up is a bit cumbersome. Let the vector $Y_t(i)$ index the ith ordering of informational release (there are $N!$ where for convenience we will assume no two firms can choose the same day to release information):

$$Y_t(i) = (y_{1t}, y_{2t}, \ldots y_{Nt}) \quad of \quad all \quad possible \quad orderings.$$

The i designates a specific ordering of information release. Corresponding to each informational ordering release is a forecast depending on the order release. To aid in description, let $N = 3$ with the following sequence of informational release sequence: $Y_t(1) = (y_{1t}, y_{2t}^1, y_{3t}^1, y_{3t}^{1:2})$ indicates that Firm 1 releases firstly, followed by Firm 2 and finally Firm 3, where the superscript 1 indicates that Firm 1 has announced its earnings and the superscript 1 : 2 indicate that Firm 1 and Firm 2 have released their earnings hereafter; $Y_t(2) = (y_{2t}, y_{1t}^2, y_{3t}^2, y_{3t}^{2:1})$ indicates that Firm 2 releases first, followed by Firm 1 and finally Firm 3; $Y_t(3) = (y_{1t}, y_{3t}^1, y_{2t}^1, y_{2t}^{1:3})$ indicates that Firm 1 releases first, followed by Firm 3 and finally Firm 2; $Y_t(4) = (y_{3t}, y_{1t}^3, y_{2t}^3, y_{2t}^{3:1})$ indicates that Firm 3 releases first, followed by Firm 1 and finally Firm 2; $Y_t(5) = (y_{2t}, y_{3t}^2, y_{1t}^2, y_{1t}^{2:3})$ indicates that Firm 2 releases first, followed by Firm 3 and finally Firm 1; $Y_t(6) = (y_{3t}, y_{2t}^3, y_{1t}^3, y_{1t}^{3:2})$ indicates that Firm 3 releases first, followed by Firm 2 and finally Firm 1. Therefore, there are six possible different orderings.

The forecast error for the first firm is zero and earnings for the remaining two firms is done by the updating formula for the VAR using the earnings information from Firm 1. Each of these produces a forecast error $(y_{2t+1} - y_{2t+1|t}^1)$ and $(y_{3t+1} - y_{3t+1|t}^1)$. Next, the second firm releases its information and the VAR forecast is calculated for updating based on Firm 1 and Firm 2 earnings release and this leaves firm 3 with a final forecast error $(y_{3t+1} - y_{3t+1|t}^{1:2})$. Thus, the sum of squared forecast errors for this ordering $Y_t(1)$ is

$$
S^{1:2} = \sum_{t=T}^{T+s} \frac{1}{s} \left\{ (y_{2t+1} - y_{2t+1|t}^1)^2 + (y_{3t+1} - y_{3t+1|t}^1)^2 + (y_{3t+1} - y_{3t+1|t}^{1:2})^2 \right\},
$$

where s denotes the number of forecasting periods. We see that Firm 3 contributes twice to the squared forecast in $S^{1:2}$. In this case there are five other possible squared forecast sequences: $S^{1:3}, S^{2:1}, S^{2:3}, S^{3:1}, S^{3:2}$. We choose the smallest S^i.

Clearly, as N gets bigger, the complexity of the search and the number of cases rises quickly. One unusual feature is the repeating of the forecast error for firms whose information has not yet been released. We could in principal discount here, but the time period is so short that this has little impact. The forecast error for the firm is repeated because there is delay in reporting their earnings. If the other firms give valuable information in forecasting this firm's earnings, this forecast error will not be large.

To conduct this exercise, we will use the entire VAR sample period to estimate all magnitudes. Each combination of real-time information is required. For the last case we need to calculate two per ordering, so that we have in total twelve. In general, there will be $(N-1) \times N!$ which with the big six Canadian banks is 3600 updates!

For $N = 3$ variables and $k = 1$ one-step-ahead forecast case, if Firm 1 releases first, the

mean squared errors for forecasting Firm 2 and Firm 3 respectively, by (10), are

$$(y_{2T+1} - y^1_{2T+1|T})^2 = (1 - \rho^2_{12})\sigma^2_2$$

and

$$(y_{3T+1} - y^1_{3T+1|T})^2 = (1 - \rho^2_{13})\sigma^2_3.$$

From (13), the mean squared error for forecasting Firm 3, knowing Firm 1 and Firm 2's current value, is denoted by $(y_{3T+1} - y^{1:2}_{3T+1|T})^2$, which is

$$(y_{3T+1} - y^{1:2}_{3T+1|T})^2 = (1 - \frac{\rho^2_{13} + \rho^2_{23} - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho^2_{12}})\sigma^2_3.$$

Now the sum of squared forecast errors for the first ordering of informational release $Y_t(1)$ is

$$
\begin{aligned}
S^{1:2} &= \sum_{t=T}^{T+s} \frac{1}{s} \left\{ (y_{2t+1} - y^1_{2t+1|t})^2 + (y_{3t+1} - y^1_{3t+1|t})^2 + (y_{3t+1} - y^{1:2}_{3t+1|t})^2 \right\} \\
&= (1 - \rho^2_{12})\sigma^2_2 + (1 - \rho^2_{13})\sigma^2_3 + (1 - \frac{\rho^2_{13} + \rho^2_{23} - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho^2_{12}})\sigma^2_3.
\end{aligned}
\tag{18}
$$

Consequently, if Firm 2 releases first, the MSE for forecasting Firm 1 and Firm 3 respectively, by (10), are

$$(y_{1T+1} - y^2_{1T+1|T})^2 = (1 - \rho^2_{12})\sigma^2_1$$

and

$$(y_{3T+1} - y^2_{3T+1|T})^2 = (1 - \rho^2_{23})\sigma^2_3.$$

From (13), the mean squared error for forecasting firm 3, knowing firm 2 and firm 1's current value, is denoted by $(y_{3T+1} - y^{2:1}_{3T+1|T})^2$, which is

$$(y_{3T+1} - y^{2:1}_{3T+1|T})^2 = (1 - \frac{\rho^2_{13} + \rho^2_{23} - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho^2_{12}})\sigma^2_3.$$

Now the sum of squared forecast errors for the second ordering of informational release $Y_t(2)$ is

$$
\begin{aligned}
S^{2:1} &= \sum_{t=T}^{T+s} \frac{1}{s} \left\{ (y_{1t+1} - y^2_{1t+1|t})^2 + (y_{3t+1} - y^2_{3t+1|t})^2 + (y_{3t+1} - y^{2:1}_{3t+1|t})^2 \right\} \\
&= (1 - \rho^2_{12})\sigma^2_1 + (1 - \rho^2_{23})\sigma^2_3 + (1 - \frac{\rho^2_{13} + \rho^2_{23} - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho^2_{12}})\sigma^2_3.
\end{aligned}
\tag{19}
$$

Likewise, suppose that Firm 1 releases, following Firm 3 and Firm 2 finally, the sum of squared forecast errors for the third ordering of informational release $Y_t(3)$ is

$$
\begin{aligned}
S^{1:3} &= \sum_{t=T}^{T+s} \frac{1}{s} \left\{ (y_{3t+1} - y_{3t+1|t}^1)^2 + (y_{2t+1} - y_{2t+1|t}^1)^2 + (y_{2t+1} - y_{2t+1|t}^{1:3})^2 \right\} \\
&= (1 - \rho_{13}^2)\sigma_3^2 + (1 - \rho_{12}^2)\sigma_2^2 + (1 - \frac{\rho_{12}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{13}^2})\sigma_2^2.
\end{aligned}
\tag{20}
$$

The sum of squared forecast errors for the fourth ordering of informational release $Y_t(4)$ is

$$
\begin{aligned}
S^{3:1} &= \sum_{t=T}^{T+s} \frac{1}{s} \left\{ (y_{1t+1} - y_{1t+1|t}^3)^2 + (y_{2t+1} - y_{2t+1|t}^3)^2 + (y_{2t+1} - y_{2t+1|t}^{3:1})^2 \right\} \\
&= (1 - \rho_{13}^2)\sigma_1^2 + (1 - \rho_{23}^2)\sigma_2^2 + (1 - \frac{\rho_{12}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{13}^2})\sigma_2^2.
\end{aligned}
\tag{21}
$$

By symmetry, the sum of squared forecast errors for fifth ordering of informational release $Y_t(5)$ is

$$
\begin{aligned}
S^{2:3} &= \sum_{t=T}^{T+s} \frac{1}{s} \left\{ (y_{3t+1} - y_{3t+1|t}^2)^2 + (y_{1t+1} - y_{1t+1|t}^2)^2 + (y_{1t+1} - y_{1t+1|t}^{2:3})^2 \right\} \\
&= (1 - \rho_{23}^2)\sigma_3^2 + (1 - \rho_{12}^2)\sigma_1^2 + (1 - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2})\sigma_1^2
\end{aligned}
\tag{22}
$$

whereas the sum of squared forecast errors for the sixth ordering of informational release $Y_t(6)$ is

$$
\begin{aligned}
S^{3:2} &= \sum_{t=T}^{T+s} \frac{1}{s} \left\{ (y_{2t+1} - y_{2t+1|t}^3)^2 + (y_{1t+1} - y_{2t+1|t}^3)^2 + (y_{1t+1} - y_{1t+1|t}^{3:2})^2 \right\} \\
&= (1 - \rho_{23}^2)\sigma_2^2 + (1 - \rho_{13}^2)\sigma_1^2 + (1 - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2})\sigma_1^2.
\end{aligned}
\tag{23}
$$

To simplify the analysis, suppose $\sigma_1 = \sigma_2 = \sigma_3 = 1$. The timing of announcement for three firms is as follows.

Proposition 6 *Suppose $|\rho_{13}| > |\rho_{12}| > |\rho_{23}|$, by minimization objective function, the informational ordering is Firm 1, followed by Firm 2, and finally Firm 3. In other words, the firm with the biggest correlation to other firms releases its earnings earlier in order to obtain the minimum mean squared forecast error.*

Proof. *See Appendix.*

This proposition shows the timing of announcement when the variances are the same for three firms. It shows that the timing of announcement is that the firm (say Firm 1), whose correlation with another firm (say Firm 3) is the largest and whose correlation with third firm (say Firm 2) is the second biggest, releases its information first.

In general, the following proposition is a general case and holds for three firms.

Proposition 7 *Suppose $\sigma_2 > \sigma_3 > \sigma_1$, $|\rho_{21}| > |\rho_{23}|$, and $|\rho_{21}| > |\rho_{31}|$. The informational ordering is Firm 2, followed by Firm 3, and finally Firm 1. In other words, the firm with the largest variance and the biggest correlation to other firms releases its information first in order to obtain the minimum mean squared forecast error.*

Proof. *See Appendix.*

The purpose of this proposition is to minimize S, that is, $min\{S^{1:2}, S^{2:1}, S^{1:3}, S^{3:1}, S^{2:3}, S^{3:2}\}$. According to the the minimum mean squared error criterion, the objection function is to have the firm that has the most volatility and is highly correlated to the least volatile firm release its information first.

3 Simulation

This section is based on Zhu (2008). Only a relevant set of results is presented here.

3.1 Bootstrap v.s Asymptotic Confidence Intervals

The first step in constructing a bootstrap DGP is to estimate the bivariate VAR (1) model (2) by maximum likelihood estimation, yielding the restricted estimates $\hat{A}$, $\hat{\Omega}_\epsilon$, where

$$\hat{A} = \begin{pmatrix} \hat{a}_{11} & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \end{pmatrix},$$

and

$$\hat{\Omega}_\epsilon = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\rho}_{12}\hat{\sigma}_1\hat{\sigma}_2 \\ \hat{\rho}_{12}\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_2^2 \end{pmatrix}.$$

Then, the bootstrap DGP is given by

$$
\begin{aligned}
y_{1t}^* &= \hat{a}_{11}y_{1t-1}^* + \hat{a}_{12}y_{2t-1}^* + \epsilon_{1t}^* \\
y_{2t}^* &= \hat{a}_{21}y_{1t-1}^* + \hat{a}_{22}y_{2t-1}^* + \epsilon_{2t}^* \quad \epsilon_t^* \sim NID(0, \hat{\Omega}_\epsilon).
\end{aligned}
$$

which is only the element of the model (2) characterized by the parameter estimates under the assumptions, with stars to indicate simulated data.

In order to draw a bootstrap sample from the bootstrap DGP, we first draw a two-vector $(\epsilon_1^*, \epsilon_2^*)'$ from the $N(0, \hat{\Omega}_\epsilon)$ distribution. The presence of a lagged dependent variable implies that the bootstrap samples must be constructed recursively. Hence, y_{1t}^* and y_{2t}^* respectively, the tth element of the bootstrap sample must depend on both y_{1t-1}^* and y_{2t-1}^* simultaneously. The recursive rule for generating a bootstrap sample is

$$
\begin{aligned}
y_{1,1}^* &= \hat{a}_{11} y_{1,0} + \hat{a}_{12} y_{2,0} + \epsilon_{1,1}^* \\
y_{2,1}^* &= \hat{a}_{21} y_{1,0} + \hat{a}_{22} y_{2,0} + \epsilon_{2,1}^* \\
y_{1,2}^* &= \hat{a}_{11} y_{1,1}^* + \hat{a}_{12} y_{2,1}^* + \epsilon_{1,2}^* \\
y_{2,2}^* &= \hat{a}_{21} y_{1,1}^* + \hat{a}_{22} y_{2,1}^* + \epsilon_{2,2}^* \\
&\quad . \qquad . \qquad . \qquad . \\
&\quad . \qquad . \qquad . \qquad . \\
y_{1t}^* &= \hat{a}_{11} y_{1t-1}^* + \hat{a}_{12} y_{2t-1}^* + \epsilon_{1,t}^* \\
y_{2t}^* &= \hat{a}_{21} y_{1t-1}^* + \hat{a}_{22} y_{2t-1}^* + \epsilon_{2,t}^*.
\end{aligned}
$$

Notice that the bootstrap sample is conditional on the observed value of $y_{1,0}$ and $y_{2,0}$, where we set them equal to zero initially.

To investigate the small sample behavior of the predictor with estimated coefficients, we consider the bootstrap by adopting a bivariate VAR model with coefficient parameters $a_{11} = 0.78$, $a_{12} = 0.1$, $a_{21} = 0.81$, and $a_{22} = 0.17$. The coefficient parameters we employed are consistent with the coefficient estimates in the application to U.S monthly labor data[4]. We generate data from above bivariate VAR (1) model with the sample size $T + 1 = 101$, where the value at time $T + 1$, $y_{1,101}$, is the real-time information we know in advance, and $y_{2,101}$, is the out-of-sample real value for forecasting comparison. To observe how the correlation coefficient ρ affects the forecasting performance in the updated bivariate VAR model, we choose $\rho = -0.7, -0.3, 0, 0.4$, and 0.8 respectively. We first compute $\hat{A}$, $\hat{\Omega}$, $\hat{fcast}$ and s_{fcast} using the original data y_1 and y_2. We then generate $B = 299$ bootstrap samples using a DGP characterized by relevant estimates, such as $\hat{A}$ and $\hat{\Omega}$. For each of the bootstrap samples, we compute the updated VAR forecast $fcast^*$ and its standard error s_{fcast}^* and its residual correlation $\hat{\rho}_{12}$. The bootstrap t-statistic is then calculated by

$$
t_j^* = \frac{fcast^* - \hat{fcast}}{\sqrt{(1 - \hat{\rho}_{12})} \, s_{fcast}^*}.
$$

Consider the $1 - \alpha$ confidence interval. Let us denote $r_{\alpha/2}$ as the smallest integer not less

---

[4]The results are presented in Gregory and Zhu (2008).

Table 1: How Correlations Affect the Small Sample Behavior of the Predictor
$Repetition = 1000$, $Bootstrap = 299$, and $Sample\ size = 100$

| | | | percent of actual values falling in the forecast interval | | |
ρ	MSE	Average S.D	90%	95%	99%
-0.7	0.51	0.747	88.5	93.7	98.9
			(87.4)	(93.1)	(97.8)
-0.3	0.91	0.966	84.9	90.7	97.2
			(89.4)	(94.1)	(97.8)
0	1	0.984	85.9	91.5	97.5
			(89.3)	(94.7)	(98.6)
0.4	0.84	0.957	86.7	91.4	97.5
			(88.0)	(92.8)	(97.8)
0.8	0.36	0.630	90.6	94.2	98.1
			(86.6)	(92.3)	(99.1)

† MSE is computed by $(1 - \rho_{12}^2)\sigma_2^2$ whereas the average standard deviation (Average S.D) is calculated by $1/rep \sum_{i=1}^{rep}(forecast - true\ value)^2$

† The number in the bracket indicates the percentage of actual values falling in the forecast interval by performing 1000 times Monte Carlo simulation.

than $\alpha B/2$. We sort the t_j^* from smallest to largest and denote by $c_{\alpha/2}^*$ the entry in the sorted list indexed by $r_{\alpha/2}$. Then the upper limit of the confidence interval is defined by $f\hat{c}ast - s_{fcast}\ c_{\alpha/2}^*$. Further, the lower limit of the confidence interval is $f\hat{c}ast - s_{fcast}\ c_{1-(\alpha/2)}^*$, where $c_{1-\alpha/2}^*$ is the entry indexed by $r_{1-(\alpha/2)}$ when the t_j^* are sorted in ascending order. Thus the asymmetric studentized bootstrap confidence interval can be written as

$$[fcast_l,\ fcast_u] = [f\hat{c}ast - \sqrt{(1 - \hat{\rho}_{12})}\ s_{fcast}\ c_{1-(\alpha/2)}^*,\ f\hat{c}ast - \sqrt{(1 - \hat{\rho}_{12})}\ s_{fcast}\ c_{\alpha/2}^*\ ].$$

We repeat each simulation $R = 1000$ times for different correlations. Table 1 reports the MSE, average standard deviation, and the numbers of actual values falling in the intervals, such as 90%, 95% and 99% confidence intervals (CI), for the bootstrap DGP.

To illustrate the asymptotic approximation of the predictor, we consider the Monte Carlo experiment through the same bivariate VAR model with the same coefficient parameters. The main difference between a bootstrap and a Monte Carlo simulation is that, for the former, it is necessary to estimate a bootstrap DGP from which to draw the simulated samples, whereas, for the latter, the DGP is assumed to be known. We generate data from the above

bivariate VAR (1) model with the sample size $T + 1 = 101$, where the value at time $T + 1$, $y_{1,101}$, is the real-time information we know in advance, whereas, $y_{2,101}$, is the out-of-sample real value for forecasting comparison. We compute the asymptotic confidence interval by

$$[fcast_l, \ fcast_u] = [\hat{fcast} - \sqrt{(1 - \hat{\rho}_{12})} \ s_{fcast} \ \sqrt{c_\alpha}, \ \hat{fcast} + \sqrt{(1 - \hat{\rho}_{12})} \ s_{fcast} \ \sqrt{c_\alpha} \ ],$$

where c_α is the asymptotic critical value. When $\alpha = 0.9, 0.95$, and 0.99, the corresponding squared root of the critical values are 1.645, 1.96, and 2.576 respectively. We repeat each simulation $R = 1000$ times for different correlations as well. The values in the bracket of Table 1 indicates the numbers of actual values falling in the intervals based on the asymptotic approximation DGP.

Note that as the positive value of the correlation coefficient ρ becomes large in both the bootstrap and the asymptotic case, the MSE of the updated bivariate VAR forecast is relatively smaller. So is the average standard deviation. Specifically, when the correlation coefficient approaches one, we have perfect linear association, and the updated bivariate VAR forecast has the best performance. This evidence is consistent with our finding of the updated bivariate VAR forecast in section 2.1; that is the updated multivariate time series forecasting will be more accurate with higher correlation coefficients among observation innovations. Further, since the sample size we choose is relatively small, the asymptotic confidence intervals are wider than the bootstrap confidence intervals. Therefore, the percentage of actual values falling in the forecast interval by the Monte Carlo simulation is larger than the one by the bootstrap method.

3.2   Timing of Announcement

In this section, we examine the finite sample performance of the release sequence using Monte Carlo simulations. We consider the case of $N = 3$. We choose the parameters which are obtained by estimating three Canadian banks'[5] quarterly earnings per share data in the sample periods of 1982Q2 to 2008Q1. We adopt simulation designs that satisfy the following conditions.

$$\begin{pmatrix} \triangle RBC_t \\ \triangle BNS_t \\ \triangle TD_t \end{pmatrix} = \begin{pmatrix} -0.46 & -0.16^* & -0.04^* \\ -0.02^* & -0.44 & 0.07^* \\ -0.13^* & 0.27^* & -0.14^* \end{pmatrix} \begin{pmatrix} \triangle RBC_{t-1} \\ \triangle BNS_{t-1} \\ \triangle TD_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{RBC,t} \\ \epsilon_{BNS,t} \\ \epsilon_{TD,t} \end{pmatrix}$$

The residual correlation coefficients are $\rho_{12} = 0.27$, $\rho_{13} = 0.69$, and $\rho_{23} = 0.08$. The

---

[5]These three Canadian banks are Royal Bank of Canada, Bank of Nova Scotia, Toronto-Dominion Banks

Table 2: Timing of Announcement

	Case I			Case II		
	Actual	Mean	S.D	Actual	Mean	S.D
$\hat{\rho}_{(RBC,BNS)}$	0.27	0.27	0.09	0.27	0.27	0.09
$\hat{\rho}_{(RBC,TD)}$	0.69	0.69	0.05	0.69	0.69	0.05
$\hat{\rho}_{(TD,BNS)}$	0.08	0.08	0.10	0.08	0.08	0.10
$\hat{\sigma}^2_{RBC}$	1	0.97	0.14	0.27	0.26	0.04
$\hat{\sigma}^2_{BNS}$	1	0.97	0.14	0.07	0.07	0.01
$\hat{\sigma}^2_{TD}$	1	0.97	0.14	0.01	0.01	0.00
RBC, BNS, TD	31.69%			0		
BNS, RBC, TD	0.74%			0		
RBC, TD, BNS	38.17%			100%		
TD, RBC, BNS	0			0		
BNS, TD, RBC	2.85%			0		
TD, BNS, RBC	26.55%			0		

† The entry in the bottom panel is the percentage of the sum of squared forecast errors. The first column in the bottom panel is the release order. The second column 'Case I' is the case where the residual variances are $\sigma_1 = \sigma_2 = \sigma_3 = 1$. The third column 'Case II' is the case where the residual variances are $\sigma_1^2 = 0.27$, $\sigma_2^2 = 0.07$, and $\sigma_3^2 = 0.01$.
† S.D denotes the standard deviation.

residual variances are $\sigma_1^2 = 0.27$, $\sigma_2^2 = 0.07$, and $\sigma_3^2 = 0.01$.

We consider two cases as in section 2.5. The first simple case is to set $\sigma_1 = \sigma_2 = \sigma_3 = 1$. The second general case is to set $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$, where we obtain the residual variances from three Canadian banks. To fit our setting, we set $\sigma_1^2 = 0.27$, $\sigma_2^2 = 0.07$, and $\sigma_3^2 = 0.01$. We generate data from the following model with sample size $T + 1 = 1001$ for both cases, where the value at time $T + 1$ is taken as either the out-of-sample real value for comparison or real-time information we know in advance.

$$
\begin{aligned}
y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + a_{13}y_{3t-1} + \epsilon_{1t}, \quad t = 1, ..., T+1 \\
y_{2t} &= a_{21}y_{1t-1} + a_{22}y_{2t-1} + a_{23}y_{3t-1} + \epsilon_{2t}, \quad t = 1, ..., T+1 \\
y_{3t} &= a_{31}y_{1t-1} + a_{32}y_{2t-1} + a_{33}y_{3t-1} + \epsilon_{3t}, \quad t = 1, ..., T+1
\end{aligned}
$$

We run 10000 experiments. The simulation results are reported in Table 2.

We set the assumptions in 'Case I' the same as Proposition 6, where $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and $\rho_{13} > \rho_{12} > \rho_{23}$. We find that there is a 32% possibility that the disclosure sequence of Firm 1, Firm 2, and Firm 3 will have the least sum of forecast squared errors. We know that sequences {1, 2, 3}, {1, 3, 2}, and {3, 2, 1} have the relatively large probability to have the

minimum sum of forecast squared errors. The result induces that correlation coefficients do play a role in determining the maximum disclosure order.

Given the assumptions $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ as well as $\rho_{13} > \rho_{12}$ and $\rho_{13} > \rho_{23}$ in 'Case II', the simulation results show that the disclosure sequence, Firm 1, Firm 3, and Firm 2, has the minimum sum of forecast squared errors by 100% of chance. In other words, based on the information release criterion - minimize standardized forecast squared errors at each informational release - the maximum disclosure order is unique (Firm 1, followed by Firm 3, and then Firm 2). These simulation results coincide with our theoretical finding in Proposition 7.

4   Application to the Canadian Banking Industry

As of January 31, 2008, the Canadian banking industry is made up of 20 domestic banks, 24 foreign bank subsidiaries, 22 full service foreign bank branches, and 7 foreign bank lending branches. In total these institutions manage over \$2.6 trillion in assets.[6] Table 3 shows that the big six banks dominate the market with 90.01% of all banking assets. Studying and forecasting the big six banks' earning per share has long been an interest of market participants and academics. Graham et al. (2005) employ a combination of field interviews and a survey instrument to contend that managers in financial institutions have a strong preference for smooth earnings. From CFOs' point of view, quarterly earnings for the same quarter the previous year, and the analyst consensus estimate are key metric benchmarks. Meeting and exceeding the benchmarks is very important to build credibility with the capital market, to maintain or increase stock prices, to improve the external reputation of the management team, and to convey future growth prospects. Therefore, we focus on the six leading Canadian banks' EPS in order to study the Canadian banking industry. The first, second, third, and fourth quarter financial results for the Canadian banking industry ended on January 31, April 30, July 31, and October 31, respectively. The quarterly and annual financial reports are heavily regulated and supervised by the Exchange Act.[7] The big six Canadian banks are all cross-listed on the New York Stock Exchange and the Toronto Stock Exchange. Recently, in an effort to provide more timely accounting information to market participants, the Ontario Securities Commission and the U.S. Securities and Exchange Commission changed the deadline for filing quarterly financial report statements of companies. Effective November 15, 2002, the filing deadline for quarterly reports were reduced gradually

---

[6]Source: Office of the Superintendent of Financial Institutions.

[7]http://en.wikipedia.org/wiki/Securities_Exchange_Act_of_1934.

Table 3: The Big Six Canadian Banks Ranked by Asset Size

| Ranking | | Name of Banks | Total Assets | Percentage of |
| World* | Canada** | | as of 31-Jan-08 | Total Assets |
2006	2008		$millions CDN	
40	1	Royal Bank of Canada	632,761	24.11%
49	2	Scotiabank Canada	449,422	17.13%
47	3	Toronto-Dominion Bank	435,153	16.58%
56	4	Bank of Montreal	376,825	14.36%
58	5	CIBC Canada	347,734	13.25%
119	6	National Bank of Canada	120,124	4.58%
		Total of all Banks in Canada	2,624,088.2	90.01%

* Top 150 world bank's ranked by asset size. Source: *The Banker*, July 2007
** Canadian bankers association. *Bank Financial Results: 2008 Fiscal Year*, April 2008

from 45 days to 35 days over three years under the Securities and Exchange Commission.[8] Effective March 30, 2004, the deadline for filing interim financial statements was reduced from 60 days to 45 days after the end of the interim period under the Ontario Securities Commission rules.[9] Therefore, each bank announces its periodic report within 35 days of the end of the quarterly period. For example, at the first quarter of 2009, TD announced its first quarter earnings on Wednesday, February 25 (which is delay of 25 days from January 31), RBC, CIBC, and BNC announced their first quarter financial statements on Thursday, February 26 (which is delay of 26 days after January 31), and BNS and BMO announced their first quarter earnings on Tuesday, March 3 (which is delay of 31 days from January 31). The particular dates and ordering of banks' financial announcements vary. This is the most important issue we deal with in the updated multivariate VAR forecasts.

4.1 Data

The big six Canadian banks' data set is collected from the I/B/E/S database through Wharton Research Data Service and the current quarterly financial statements of each bank. We choose the past 23 years of quarterly banks' earnings per share and individual analysts' forecast of EPS data dating from 1986Q2 to 2009Q1. We take the average of individual analysts' forecasts of EPS, and get the consensus forecast of EPS. Table 4 reports the descriptive statistics. Augmented Dickey-Fuller (ADF) nonstationary tests have been conducted, and the presence of a unit root is rejected. All six series are stationary in first differences. Since

---

[8]See Securities and Exchange Commission Release No. 33-8128: "Acceleration of Periodic Report filing Dates and Disclosure Concerning Web Access to Reports".

[9]See Ontario Securities Commission National Instrument 51-102: "Continuous Disclosure Obligations".

the test is known to have low power, even a slight rejection means that the existence of a unit root is unlikely. The time series plot of the data is provided in Figure 1. After the first difference of each time series, the plot of the quarterly changes of EPS is provided in Figure 2.

Table 4: Descriptive Statistics

Source	I/B/E/S					
Frequency	Quarterly					
Sample period	1986Q2 - 2009Q1					
Sample Size	92					
	rbc	bns	td	bmo	$cibc$	bnc
Mean	0.0013	0.0074	0.0109	0.0014	-0.0001	0.0104
S.D.	0.58	0.12	0.29	0.38	0.85	0.52
β_1	-1.48	-1.58	-1.37	-1.48	-1.08	-1.47
ADF	-16.06	-15.98	-13.77	-15.76	-10.30	-15.75
H_0	Reject at 95% (critical value = -1.95)					

[†] The coefficient β designates the autocorrelation of the series at lag i. The augmented Dickey-Fuller test is based on the following regression: $y_t - y_{t-1} = \beta_0 + \beta_1 y_{t-1} + \beta_2 \Delta y_{t-1} + \beta_3 \Delta y_{t-2} + \beta_4 \Delta y_{t-3} + \beta_5 \Delta y_{t-4} + \mu_t$. Terms added until additional lags provide no new information significant at the 5% level.
[†] S.D. denotes the standard deviation.
[†] H_0, the null hypothesis of the ADF test, is nonstationary.

Figure 1 shows that banks' profits move up and down together over time, except in 2005Q3, where there is a dramatic drop for CIBC. This is a result of the settlements reached during the quarter with U.S. regulators relating to financing and brokerage services CIBC provided to hedge funds engaged in mutual fund market timing. Settlements on two Enron-related litigation matters were negotiated in late July and finalized on August 2, 2005. These settlements have a significant impact on CIBC's financial performance for 2005. In the first quarter of 2008 there are two significant unexpected results. One is for CIBC whose profits drop dramatically.[10] The other unexpected result is for BMO. Excluding significant items, net income of BMO is $617 million. However, reported profits including significant items[11] are 255 million.

---

[10] "Our losses related to the U.S. residential mortgage market are a significant disappointment and are not aligned with our strategic imperative of consistent and sustainable performance," said Gerald T. McCaughey, President and Chief Executive Officer. "Our focus is to get CIBC back on the strategic track we set for the organization which has, for the past two years, resulted in significant value for our shareholders."

[11] Significant items are charges for certain trading activities and valuation adjustments and an increase in the general provision for credit losses.
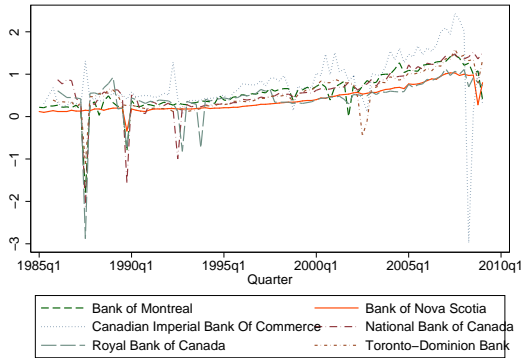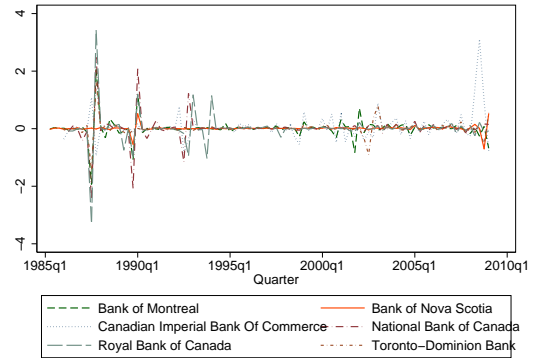
Figure 1: EPS in Level	Figure 2: Changes in EPS

Source: I/B/E/S database through Wharton Research Data Service and the current quarterly financial statements from each bank.

4.2 Forecasting

A reduced form VAR expresses each variable as a linear function of its own past values, the past values of all other variables being considered. It also captures a serially correlated error term across equations. The error terms in these regressions are the co-movements in the variables after taking past values into account. Thus, in this study the VAR involves six equations: current rbc[12] as a function of past values of the rbc, the bns, the td, the bmo, the $cibc$, and the bnc, current bns as a function of past values of rbc, the bns, the td, the bmo, the $cibc$, and the bnc, and so on. Each equation can be estimated by ordinary least squares regression. This OLS estimator is as efficient as the maximum likelihood estimator and the general least squared estimator. The number of lagged values to include in each equation is determined by Schwarz's Bayesian information criterion (SBIC), the Hannan and Quinn information criterion (HQIC), Akaike's information criterion (AIC), and the final prediction error (FPE) criterion. The latter two criteria indicate that the optimal lag selection is three, whereas the former two criteria, SBIC and HQIC, suggest that the optimal lag selection is one. For comparison we employ one lag in all time series models such as the updated VAR and the ordinary VAR. The estimated banks' EPS VAR (1) model is as follows.

_____

[12]Each of the upper-case letter indicates the abbreviate name of the big six Canadian banks, whereas each of the lower-case letter indicates time series variable of the banks

$$rbc_t = -0.55rbc_{t-1} - 0.19bns_{t-1} - 0.01td_{t-1} - 0.12bmo_{t-1} + 0.05cibc_{t-1} + 0.23bnc_{t-1}$$
$$\quad\;\; (0.18) \qquad (0.52) \qquad (0.27) \qquad (0.29) \qquad (0.06) \qquad (0.19)$$

$$bns_t = -0.05rbc_{t-1} - 0.59bns_{t-1} + 0.03td_{t-1} + 0.09bmo_{t-1} - 0.03cibc_{t-1} - 0.004bnc_{t-1}$$
$$\quad\;\; (0.04) \qquad (0.10) \qquad (0.05) \qquad (0.06) \qquad (0.01) \qquad (0.04)$$

$$td_t = -0.14rbc_{t-1} - 0.17bns_{t-1} - 0.23td_{t-1} + 0.13bmo_{t-1} + 0.01cibc_{t-1} - 0.03bnc_{t-1}$$
$$\quad\;\; (0.09) \qquad (0.27) \qquad (0.14) \qquad (0.15) \qquad (0.03) \qquad (0.10)$$

$$bmo_t = 0.01rbc_{t-1} + 0.40bns_{t-1} + 0.11td_{t-1} - 0.49bmo_{t-1} - 0.02cibc_{t-1} - 0.08bnc_{t-1}$$
$$\quad\;\; (0.12) \qquad (0.35) \qquad (0.18) \qquad (0.19) \qquad (0.04) \qquad (0.12)$$

$$cibc_t = -0.12rbc_{t-1} + 0.99bns_{t-1} + 0.79td_{t-1} - 0.41bmo_{t-1} - 0.07cibc_{t-1} + 0.14bnc_{t-1}$$
$$\quad\;\; (0.29) \qquad (0.85) \qquad (0.45) \qquad (0.46) \qquad (0.10) \qquad (0.31)$$

$$bnc_t = -0.15rbc_{t-1} - 0.31bns_{t-1} + 0.08td_{t-1} + 0.06bmo_{t-1} - 0.002cibc_{t-1} - 0.38bnc_{t-1}$$
$$\quad\;\; (0.16) \qquad (0.47) \qquad (0.25) \qquad (0.26) \qquad (0.06) \qquad (0.17) \qquad (24)$$

After fitting a VAR(1) model in (24), we implement a lagrange-multiplier (LM) test for autocorrelation in the residuals of the VAR(1) model, which is presented in table 5. Since we reject the null hypothesis that there is no autocorrelation in the residuals for any of the five orders tested, this test indicates that autocorrelation of the residuals exists.

Table 5: Residual Autocorrelation Test

lag	chi_2	df	$Prob > chi_2$
1	86.01	36	0.00
2	111.70	36	0.00
3	119.58	36	0.00
4	58.33	36	0.01
5	42.49	36	0.21

[†] H_0: no autocorrelation at lag order.
[†] Since we reject the null hypothesis that there is no autocorrelation in the residuals for all five orders tested, this test indicates that autocorrelation of the residuals exists.

In the ordinary VAR forecast, the coefficient estimates and the variance covariance matrix estimates are used to calculate the forecast mean squared error. In the updated VAR forecast, in addition to coefficient estimates and variance covariance matrix estimates, we also consider the correlation of residuals of VAR to compute the more efficient forecast mean squared

Table 6: Correlations of Residuals

	RBC	BNS	TD	BMO	CIBC	BNC
Correlation Matrix						
RBC	1.00					
BNS	0.21	1.00				
TD	0.67	0.23	1.00			
BMO	0.84	0.27	0.58	1.00		
CIBC	-0.09	0.12	-0.10	-0.08	1.00	
BNC	0.76	0.37	0.54	0.79	-0.07	1.00
Covariance Matrix						
RBC	0.25					
BNS	0.01	0.01				
TD	0.09	0.01	0.07			
BMO	0.14	0.01	0.05	0.11		
CIBC	-0.04	0.01	-0.02	-0.02	0.67	
BNC	0.18	0.02	0.06	0.12	-0.03	0.21

error. The correlation of residuals of the multivariate VAR is shown in Table 6, indicating a positive correlation between all banks, except that CIBC has negative correlations with the other three banks.

The big six banks' EPS are covariance stationary if their first two moments exist and are independent of time. Table 7 indicates that all the eigenvalues lie inside the unit circle. Therefore, the estimates satisfy the stability condition. In other words, the big six banks' EPS times series follow a VAR(1) stationary process.

Table 7: Eigenvalue Stability Condition

Eigenvalue	Modulus
-0.58	0.58
-0.52 + 0.08i	0.53
-0.52 - 0.08i	0.53
-0.45	0.45
-0.12 + 0.13i	0.17
-0.12 - 0.13i	0.17

All the eigenvalues lie inside the unit circle. VAR satisfies stability condition.

Further, a test for normally distributed disturbances after fitting VAR (1) of model (24) is conducted, for each equation and for all equations jointly. Table 8 reports a series of

statistics - a skewness statistic, a kurtosis statistic, and the Jarque-Bera statistic - against the null hypothesis that the disturbances in a VAR are normally distributed. The Jarque-Bera statistic is a combination of the other two statistics. Neither the single equation Jarque-Bera statistics nor the joint Jarque-Bera statistics come close to rejecting the null hypothesis. The result indicates that the six disturbances follow a six-dimensional multivariate normal distribution.

Table 8: Test for Normally Distributed Disturbances after VAR

Equation	Skewness/ Kurtosis	chi_2	df	$Prob > chi_2$
Jarque-Bera test				
rbc		101.09	2	0.00
bns		1470.22	2	0.00
td		244.48	2	0.00
bmo		1033.71	2	0.00
cibc		7529.49	2	0.00
bnc		185.73	2	0.00
ALL		11000	12	0.00
Skewness test				
rbc	-1.05	16.65	1	0.00
bns	-3.27	162.32	1	0.00
td	-0.58	5.14	1	0.02
bmo	-2.02	62.09	1	0.00
cibc	-4.84	355.88	1	0.00
bnc	-0.76	8.67	1	0.00
ALL		610.75	6	0.00
Kurtosis test				
rbc	7.72	84.44	1	0.00
bns	21.57	1307.90	1	0.00
td	10.95	239.34	1	0.00
bmo	19.01	971.63	1	0.00
cibc	46.50	7173.61	1	0.00
bnc	9.83	177.06	1	0.00
ALL		9953.97	6	0.00

After lag-order selection statistics for the VAR, lagrange-multiplier statistics for residual autocorrelation, and a test for normally distributed disturbances after VAR, we confirm that the big six banks' EPS data set satisfies the assumptions for performing the updated

VAR forecast. Since all the banks announce their quarterly earnings at different dates, the theoretical framework in section 2 is an accurate method for a real-time forecast.

We compute forecasts of the model over the period from 2007Q1 to 2009Q1. More precisely, we first use data from 1986Q2 through 2006Q4 to estimate the model. Once estimated, the model is used to produce forecasts one quarter ahead; i.e., forecasts for 2007Q1. Next, we add the 2007Q1 data point to the original data set, re-estimate the model, and forecast again one quarter ahead; i.e., for 2007Q2. We update our estimation and forecasts in such increments until the next-to-last data point (2008Q4) has been added to the estimation. We thus obtain series for one-quarter-ahead forecasts running from 2007Q1 through 2009Q1. These forecasts can then be confronted with observed data.

We report the results of using three different methods of forecasting the big six banks' quarterly EPS. The ordinary VAR forecasting method is to first fit the VAR (1) model in (24), then to compute the dynamic forecasts of the dependent variables in the VAR. A consensus forecast is a simple arithmetic average of all of the individual predictions collected by I/E/B/S. The updated VAR forecasting method takes announcement timing into account and conducts real-time forecasting using real-time data.The detailed estimation methodology is investigated in the following two-stage one-step-ahead VAR forecast. In the first stage, we estimate the multivariate VAR over the period from 1986Q2 through 2006Q4. Following the maximum likelihood estimation by VAR, the one-step ahead residual prediction of all six series is straightforward. In the second stage, we regress the residuals of the unknown series on the residuals of the known one. Since we observe the actual error term of the known series in 2007Q1, the best fitted residual of the unknown series in 2007Q1 is the estimated coefficient multiplied by the actual error term of the known series in 2007Q1.

Consequently, we re-estimate the multivariate VAR over the period from 1986Q2 through 2007Q1. One-step-ahead fitted values of six series are predicted. We follow the latest order of the big six banks' release for the second quarter of 2007. We regress the residuals of unknown series on the residuals of the known one and calculate the fitted residual of the unknown values of the series by multiplying the estimated coefficients in the residual regress by the actual error term of the known values of the series in 2007Q2. This represents our best fitted values of all *rbc*, *bns*, *td*, *bmo*, *cibc*, and *bnc* in 2007Q2. We can continue to forecast the big six banks' EPS once each of their latest earnings announcements is available. This dynamic forecast is based on the real time information, that is, the latest actual value of each series released for public use in advance. As well, all the six time series are relatively correlated with the others, thus this two-stage one-step-ahead VAR forecast benefits from accurate measurements.

We consider forecast periods from 2007Q1 through 2009Q1, which is nine observations.

The first forecasting observation is on the first quarter of 2007. The order of the information release at that time is TD, (BMO, BNC, and CIBC), RBC, and BNS, where the banks in parentheses indicate the same date of earnings announcements. Hence, on the first quarter of 2007, TD announces its earnings first, followed by (BMO, BNC, and CIBC) releasing at the same date. Next, RBC announces its earnings on the following date. Finally, BNS announces its earnings. So in this case we have three rounds. At the first round, given the earnings announcement of TD in 2007Q1, we can forecast the earnings for the same quarter of each of (BMO, BNC, and CIBC). At the second round, given any combinations of the earnings announcements of TD, BMO, BNC, and CIBC in 2007Q1, we can forecast the earnings for the same quarter of RBC. At the third round, we can forecast the earnings for the same quarter of BNS based on any combinations of earnings announcements of TD, BMO, BNC, CIBC, and RBC in 2007Q1.

The second forecasting observation is for the second quarter of 2007. The order of the information release at that time is BMO, TD, RBC, BNS, and (BNC and CIBC). Hence, in the second quarter of 2007, BMO announces its earnings first, followed by TD, RBC, and BNS, respectively. Finally, (BNC and CIBC) announce their earnings at the same date. So in this case we have four rounds. At the first round, given the earnings announcement of BMO in 2007Q2, we can forecast the earnings of TD for the same quarter. At the second round, given the earnings announcements of BMO and TD in 2007Q2, we can forecast the earnings of RBC for the same quarter. At the third round, we can forecast the earnings of BNS for the same quarter based on any combinations of the earnings announcements of BMO, TD, and RBC in 2007Q2. At the fourth round, we can forecast the earnings of each of (BNC and CIBC) for the same quarter based on any combinations of the earnings announcements of BMO, TD, RBC, and BNS in 2007Q2.

The third, fourth, fifth, sixth, seventh, eighth, and ninth forecasting observations follow the same pattern. The order for the information release for the third observation is TD, RBC, (BNS and BMO), and (CIBC and BNC); for the fourth observation period, the order is BMO, (TD and BNC), RBC, and (BNS and CIBC); for the fifth observation period, the order is (TD, CIBC, and BNC), RBC, and (BNS and BMO); for the sixth observation period, the order is (BNS and BMO), TD, and (RBC, CIBC, and BNC); for the seventh observation period, the order is (BNS and BMO), CIBC, and (RBC, TD, and BNC); for the eighth observation period, the order is BMO, BNS, (TD, CIBC, and BNC), and RBC; for the ninth observation period, the order is TD, (RBC, CIBC, and BNC), and (BNS and BMO). Figure 3 shows earnings forecasts of the big six Canadian banks from the first quarter of 2007 to the first quarter of 2009.

Figure 3: Earnings Forecasts of the Canadian Banks
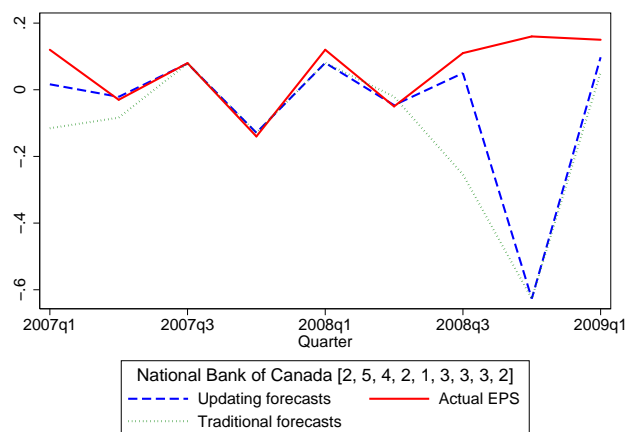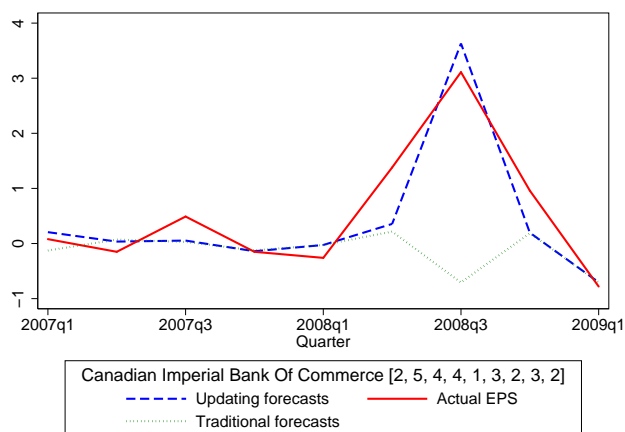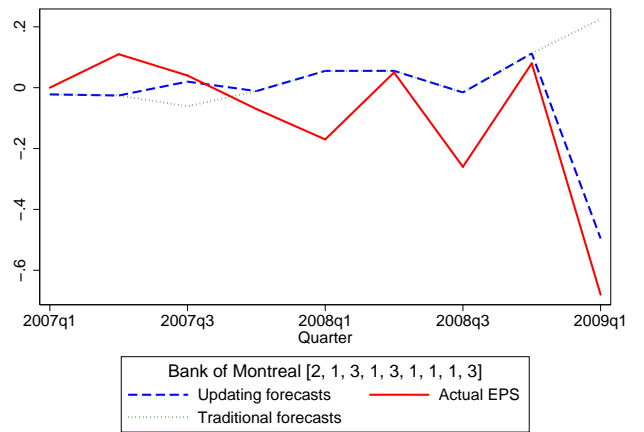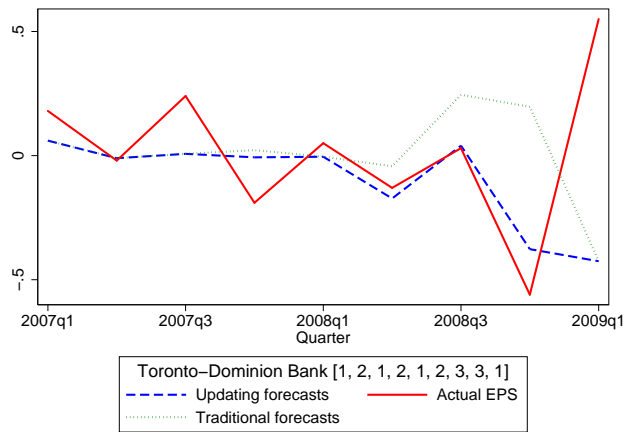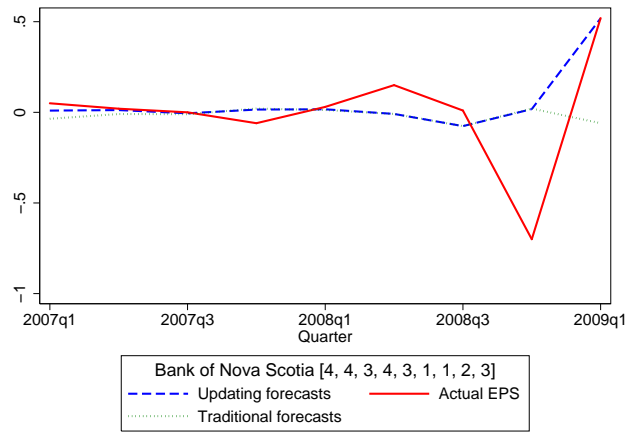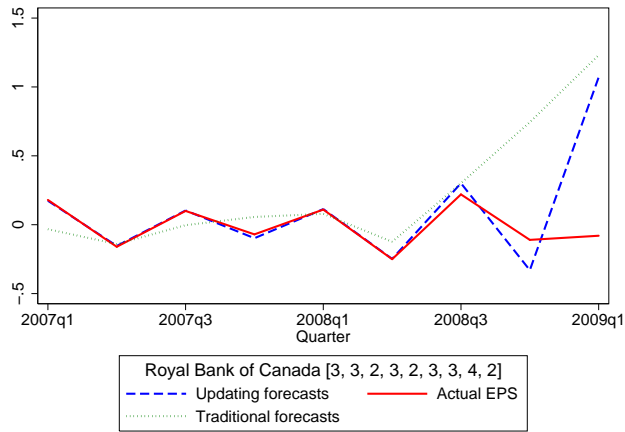
4.3 Forecast Results By Various Methods

There are numerous ways of measuring forecast accuracy. We use the forecast root mean squared error (RMSE), which for each bank can be expressed as

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^{k} (A_i - F_i)^2},$$

where k is the number of forecasting observations, A_i is actual reported value of quarterly EPSs for the forecasting observation i, and F_i is a forecast of A_i. Table 9 reports the root mean squared errors for each six banks by various methods: the updated VAR, the ordinary VAR, and consensus forecasts. The third column of Table 9 represents the efficiency gains of the updated VAR relative to the ordinary VAR whereas the fifth column of Table 9 demonstrates the efficiency gains of the updated VAR compared with consensus forecasts. We find that the RMSEs for each bank using the updated VAR forecast are smaller than those using the ordinary VAR forecast. The maximum efficiency gain of the six banks is 64% by CIBC. Although CIBC has relatively lower correlation with the rest of the five banks, two facts - CIBC earnings releases are mostly based on the rest of the five banks' announcements and CIBC has the highest variance - explain CIBC benefiting from the updated VAR forecast. This implies that banks can be well predicted by the updated VAR methodology. By using the updated VAR forecast, we find that the RMSEs of RBC, BNS, BMO, and CIBC are smaller than the RMSEs of consensus forecasts respectively. The corresponding efficiency gain for the four banks is 3%, 14%, 46% and 58%, respectively.

Table 9: Forecast Efficiency Gains

	Updating	Ordinary	Gains	Consensus	Gains
RBC	0.39	0.53	26%	0.40	3%
BNS	0.25	0.32	22%	0.29	14%
TD	0.35	0.43	19%	0.28	-20%
BMO	0.14	0.33	58%	0.26	46%
CIBC	0.49	1.37	64%	1.16	58%
BNC	0.27	0.30	10%	0.12	-56%

4.4 Thought Experiments

The big six banks' announcement dates vary from quarter to quarter. According to the covariance matrix estimation based on data from 1986Q2 to 2008Q1, the residual variance

Table 10: Reordering of Announcement Interval

	VAR	Updating VAR				
	$RMSE$	$RMSE^1$	$RMSE^2$	$RMSE^3$	$RMSE^4$	$RMSE^5$
RBC	0.27	0.03	0.02	**0.01**	0.03	0.06
BNS	0.12	0.04	0.03	0.03	**0.02**	0.03
TD	0.34	0.11	0.10	**0.09**	0.10	0.16
BMO	0.29	**0.10**	0.11	0.11	0.12	0.14
CIBC	0.60	0.26	0.26	0.24	0.26	**0.23**
BNC	0.24	0.02	**0.01**	0.03	0.04	0.12

Thought experiments are based on five observations (from 2007Q1 to 2008Q1). The entries in the table are the forecast root mean squared errors (RMSE).

ordering from largest to least is RBC, BNC, BMO, TD, CIBC, and BNS. Based on Proposition 7, we have the minimum mean squared forecast error when the most volatile firm releases its information earliest. We consider keeping this strategy disclosure as an optimal sequence in comparison with the actual release ordering. The thought experiments are studied to determine whether there exists the maximum disclosure.

We consider the forecast period of 2007Q1 through 2008Q1, consisting of five one-step-ahead forecast observations. The first column in Table 10 is the RMSEs using the ordinary VAR forecast. The rest of the five columns is the RMSEs using the updated VAR forecast. Clearly, all the RMSEs by using the updated VAR forecast are smaller than the one calculated by the ordinary VAR forecast. The superscript represents the sequence of information releases. For instance, the column $RMSE^1$ represents that the updated VAR forecast is based on information release of any one of the big six Canadian banks, whereas the column $RMSE^2$ denotes that the updated VAR forecast is based on information release of any two combinations of the big six Canadian banks, and so on. The row demonstrates the RMSEs by using the ordinary VAR forecast, the minimum RMSE by the updated VAR forecast with information release of any one of the rest banks, the minimum RMSE with information release of any two combinations of the rest banks, and so on.

In addition, the bold numbers illustrate the minimum RMSE for each bank when this bank's earnings forecast is based on any combinations of the release sequence of the rest banks. For instance, the minimum RMSE of BMO is 0.10, which is the minimum RMSE of BMO based on information release of any one of the rest banks by using the updated VAR forecast method.

Based on this reordering of announcement interval exercise, BMO benefits from any one of the rest banks' information release in advance while CIBC benefits from any four

combinations of the rest five banks' information release in advance. As a result, the timing of announcement to obtain the maximum disclosure is to have BMO release first, followed by BNC, then are (TD and RBC), then BNS, and finally CIBC.

5 Conclusions

Empirical studies demonstrate that most macroeconomic or financial time series do not end at a common point. When we make multivariate forecasts, we usually ignore the latest real-time information at hand and adopt the same ending period.

In this paper, we derive a practical method to update forecasts in a multivariate VAR model. The theoretical framework shows that the current known observations of one variable are always going to be useful for forecasting the current unknown observations of other variables. Therefore, a higher correlation among observation innovations of multiple variables implies that the mean squared forecast error of the current unknown observations of other variables will be accurate for a longer period.

We first derive the general mean squared error of a multistep-ahead forecast in an updated bivariate VAR, which is smaller than the corresponding standard VAR forecast. The theoretical model is estimated using U.S. nonfarm payroll employment data[13] and broadly replicates many of the features of the updated bivariate VAR forecast. The estimated model is then used to perform a long-horizon forecast to assess the forecast mean squared errors. Moreover, we study the updated bivariate VAR forecast with data from two additional periods known in advance. The example of Canada Telehealth, which uses data as a forecast to alert physicians to flu outbreak or vaccine failure, also indicates in practice that this updated bivariate VAR forecast can be applied broadly. We find that it is more accurate to use the updated bivariate VAR forecast given a certain sufficient condition.

In addition, we analyze the general multivariate MSE of a one-step-ahead forecast in an updated multivariate VAR forecast. We show that as more information is available, the mean squared forecast error is reduced. Studying asymptotic and exact small sample performances shows that changes to the correlation coefficient do affect forecast errors. The updated forecasts more accurately match the real data by considering the real-time information compared to the ordinary VAR forecasts. The timing issues based on real-time multiple variables are discussed.

This updated VAR forecast has direct application in macroeconomic and financial time series. More specifically, we study the earnings forecast accuracy for the big six Canadian banks. This paper examines whether one bank's earnings announced at the current quarter

---

[13]Gregory and Zhu (2008).

helps to forecast the forthcoming quarterly earnings of the rest of the banks. By using the updated VAR forecast, we find that the relative efficiency gain is 33% in the one-step-ahead forecast compared to the ordinary VAR forecast, and 7% compared to consensus forecasts. Further, studying alternative orderings of earnings release for the six banks helps to explore the most informative release. The objective function we address in this paper is to minimize mean squared forecast errors. Thought experiments suggest that there are substantial efficiency gains due to maximizing disclosure criterion. We find that if banks' order of information release were to change, forecast errors could be substantially reduced. These experiments emphasize that evaluating the release ordering is crucial in determining forecast accuracy.

One important characteristic of capital markets is that the timing of a quarterly or annual report announcement over the same time period of observation is often spread over weeks. Another characteristic is that different firms in the same industry sector are relatively correlated. Taking into account these two characteristics, we note that our updated multivariate VAR forecast method has broad applicability. Future research will extend applications to earnings forecasts for firms in different industry sectors and to forecasting macroeconomic indicators.

Appendix

Proof of proposition 1. Given the full information set $\mathcal{I}_1$, we stand at $T+1$. We know the time series of y_1 from 1 through $T+1$ while we only know the time series of y_2 from 1 through T.

For the forecast horizon at time $T+1$, we observe ϵ_{1T+1}, since $\epsilon_{1T+1} = y_{1T+1} - (a_{11}y_{1T} + a_{12}y_{2T})$. If we regress ϵ_2 on ϵ_1, the conditional expectation

$$
\begin{aligned}
E[\epsilon_{2T+1}|\epsilon_{1T+1}] &= \frac{Cov(\epsilon_2, \epsilon_1)}{Var(\epsilon_1)}\epsilon_{1T+1} \\
&= \frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2}\epsilon_{1T+1}.
\end{aligned}
$$

So we can forecast the residual $\hat{\epsilon}_{2T+1}$ by the relationship $\hat{\epsilon}_{2T+1} = \epsilon_{2T}(1) = (\rho_{12}\sigma_2/\sigma_1)\epsilon_{1T+1}$. Then forecast error becomes:

$$
\begin{aligned}
y_{2T+1} - \hat{y}_{2T+1} &= \epsilon_{2T+1} - \hat{\epsilon}_{2T+1} \\
&= \epsilon_{2T+1} - \rho_{12}\frac{\sigma_2}{\sigma_1}\epsilon_{1T+1}
\end{aligned}
$$

The variance of the forecast error of y_2 at $T+1$ is

$$
\begin{aligned}
MSE^u[y_{2T}(1)] &= Var[y_{2T+1} - \hat{y}_{2T+1}] \\
&= Var[\epsilon_{2T} - \epsilon_{2T}(1)] \\
&= Var[\epsilon_{2T+1} - \rho_{12}\frac{\sigma_2}{\sigma_1}\epsilon_{1T+1}] \\
&= (1 - \rho_{12}^2)\sigma_2^2
\end{aligned}
$$

∎

Proof of proposition 2. Given the full information set $\mathcal{I}_1$, we stand at $T+1$. We know the time series of y_1 from 1 through $T+1$ while we only know the time series of y_2 from 1 through T.

The bivariate VAR (1) can be represented by the matrices in model (1). We can obtain

the moving average model notation by iterating forward.

$$
\begin{aligned}
Y_1 &= AY_0 + \epsilon_1 \\
Y_2 &= AY_1 + \epsilon_2 = A^2 Y_0 + A\epsilon_1 + \epsilon_2 \\
&\quad \dots \\
Y_t &= A^t Y_0 + \sum_{i=0}^{t-1} A^i \epsilon_{t-i}
\end{aligned}
$$

The k-step ahead

$$
Y_{t+k} = A^k Y_t + \sum_{i=0}^{k-1} A^i \epsilon_{t+k-i}
$$

Given ϵ_{t+j}, for $j > 0$, is uncorrelated with y_{t-i}, for $i \geq 0$, the minimal forecast MSE by the ordinary VAR is

$$
\begin{aligned}
MSE[Y_t(k)] &= E\left(\sum_{i=0}^{k-1} A^i \epsilon_{t+k-i}\right)\left(\sum_{i=0}^{k-1} A^i \epsilon_{t+k-i}\right)' \\
&= \sum_{i=0}^{k-1} A^i \Omega_\epsilon A^{i'},
\end{aligned}
\tag{25}
$$

At $T+1$, the ordinary bivariate VAR forecast MSE by equation (25) is

$$
\begin{aligned}
MSE[Y_T(1)] &= \Omega_\epsilon \\
&= \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.
\end{aligned}
$$

For the updated VAR, we adopt all the time periods from 1 through $T+1$ for y_1 and the time periods from 1 through T for y_2. Since ϵ_1 and ϵ_2 are correlated, as well, the innovation ϵ_{1T+1} is known. By a linear regression, the best predictor of ϵ_{2T+1} based on ϵ_{1T+1} is $(\rho_{12}\sigma_2/\sigma_1)\epsilon_{1T+1}$. Thus the forecast error is

$$
\begin{aligned}
Y_{T+1} - \hat{Y}_{T+1} &= \begin{pmatrix} y_{1T+1} - \hat{y}_{1T+1} \\ y_{2T+1} - \hat{y}_{2T+1} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \epsilon_{2T+1} - \hat{\epsilon}_{2T+1} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \epsilon_{2T+1} - \rho_{12}\frac{\sigma_2}{\sigma_1}\epsilon_{1T+1} \end{pmatrix}
\end{aligned}
$$

and the MSE or forecast error covariance matrix of the updated bivariate VAR is

$$
\begin{aligned}
MSE^u[Y_T(1)] &= Var[Y_{T+1} - \hat{Y}_{T+1}] \\
&= E \begin{pmatrix} \epsilon_{1T+1} - \hat{\epsilon}_{1T+1} \\ \epsilon_{2T+1} - \hat{\epsilon}_{2T+1} \end{pmatrix} \begin{pmatrix} \epsilon_{1T+1} - \hat{\epsilon}_{1T+1} \\ \epsilon_{2T+1} - \hat{\epsilon}_{2T+1} \end{pmatrix}' \\
&= \begin{pmatrix} 0 & 0 \\ 0 & (1 - \rho_{12}^2)\sigma_2^2 \end{pmatrix}.
\end{aligned}
$$

At $T + 2$,

$$
\begin{aligned}
Y_{T+2} &= AY_T + \epsilon_{T+2} + A\epsilon_{T+1} \\
\hat{Y}_{T+2} &= AY_T + \hat{\epsilon}_{T+2} + A\hat{\epsilon}_{T+1}.
\end{aligned}
$$

The ordinary bivariate VAR forecast MSE followed by equation (25) is

$$
\begin{aligned}
MSE[Y_T(2)] &= \Omega_\epsilon + A\Omega_\epsilon A' \\
&= MSE[Y_T(1)] + A\Omega_\epsilon A'.
\end{aligned}
$$

The updated bivariate VAR forecast MSE is

$$
\begin{aligned}
MSE^u[Y_T(2)] &= E\left((\epsilon_{T+2} - \hat{\epsilon}_{T+2}) + A(\epsilon_{T+1} - \hat{\epsilon}_{T+1}) \right)\left((\epsilon_{T+2} - \hat{\epsilon}_{T+2}) + A(\epsilon_{T+1} - \hat{\epsilon}_{T+1}) \right)' \\
&= \Omega_\epsilon + A \begin{pmatrix} 0 & 0 \\ 0 & (1 - \rho_{12}^2)\sigma_2^2 \end{pmatrix} A' \\
&= \Omega_\epsilon + A \ \ MSE^u[Y_T(1)] \ \ A'.
\end{aligned}
$$

At $T + 3$,

$$
\begin{aligned}
Y_{T+3} &= AY_T + \epsilon_{T+3} + A\epsilon_{T+2} + A^2\epsilon_{T+1} \\
\hat{Y}_{T+3} &= AY_T + \epsilon_{T+3} + A\hat{\epsilon}_{T+2} + A^2\hat{\epsilon}_{T+1}.
\end{aligned}
$$

The ordinary bivariate VAR forecast MSE followed by equation (25) is

$$
\begin{aligned}
MSE[Y_T(3)] &= \Omega_\epsilon + A\Omega_\epsilon A' + A^2\Omega_\epsilon A^{2'} \\
&= MSE[Y_T(2)] + A^2\Omega_\epsilon A^{2'}.
\end{aligned}
$$

The updated bivariate VAR forecast MSE is

$$
\begin{aligned}
MSE^u[Y_T(3)] &= E\left(\left((\epsilon_{T+3} - \hat{\epsilon}_{T+3}) + A(\epsilon_{T+2} - \hat{\epsilon}_{T+2}) + A^2(\epsilon_{T+1} - \hat{\epsilon}_{T+1})\right)\right. \\
&\qquad \left.\left((\epsilon_{T+3} - \hat{\epsilon}_{T+3}) + A(\epsilon_{T+2} - \hat{\epsilon}_{T+2}) + A^2(\epsilon_{T+1} - \hat{\epsilon}_{T+1})\right)'\right) \\
&= \Omega_\epsilon + A\Omega_\epsilon A' + A^2 \begin{pmatrix} 0 & 0 \\ 0 & (1 - \rho_{12}^2)\sigma_2^2 \end{pmatrix} A^{2'} \\
&= \Omega_\epsilon + A\ MSE^u[Y_T(2)]\ A'.
\end{aligned}
$$

Recursively, the k-step ahead updating forecast error covariance matrix becomes

$$
\begin{aligned}
MSE^u[Y_T(k)] &= \sum_{i=0}^{k-2} A^i \Omega_\epsilon A^{i'} + A^{k-1} \begin{pmatrix} 0 & 0 \\ 0 & (1 - \rho_{12}^2)\sigma_2^2 \end{pmatrix} A^{k-1'} \\
&= \Omega_\epsilon + A\ MSE^u[Y_T(k-1)]\ A', \quad k \geq 2.
\end{aligned}
$$

■

Proof of proposition 3. Given the full information set $\mathcal{I}_2$, we know the time series of y_1 from 1 through $T + s$ while we only know the time series of y_2 from 1 through T.

At $T + 2$, equation (7) gives

$$
\epsilon_{1T+2} = y_{1T+2} - (a_{11}y_{1T+1} + a_{12}y_{2T+1}).
$$

Since we do not observe y_{2T+1}, we do not observe ϵ_{1T+2}. There are two ways to make a prediction on ϵ_{1T+2}. One way is to set $E(\epsilon_{1T+2}) = 0$ and to set the variance of ϵ_{1T+2} be the first element of the MSE of the ordinary bivariate VAR forecast, that is, the first element of the matrix $\Omega_\epsilon + A\Omega_\epsilon A'$. The alternative way is to predict $\hat{\epsilon}_{1T+2}$ through the residual form $\hat{\epsilon}_{1T+2} = y_{1T+2} - (a_{11}y_{1T+1} + a_{12}\hat{y}_{2T+1})$. In the latter case, the variance of the difference in error becomes
$$
Var[\epsilon_{1T+2} - \hat{\epsilon}_{1T+2}] = a_{12}^2(1 - \rho_{12}^2)\sigma_2^2.
$$

If the sufficient condition $\{a_{12}^2(1 - \rho_{12}^2)\sigma_2^2 <$ the first element of matrix $\Omega_\epsilon + A\Omega_\epsilon A'\}$ holds, we would use $\hat{\epsilon}_{1T+2}$ rather than $E(\epsilon_{1T+2}) = 0$. Then notify that

$$
\begin{aligned}
y_{2T+2} &= a_{21}y_{1T+1} + a_{22}y_{2T+1} + \epsilon_{2T+2} \\
\hat{y}_{2T+2} &= a_{21}y_{1T+1} + a_{22}\hat{y}_{2T+1} + \hat{\epsilon}_{2T+2}.
\end{aligned}
$$

By equation (3), the variance of the forecast error is followed by

$$
\begin{aligned}
Var[y_{2T+2} - \hat{y}_{2T+2}] &= a_{22}^2 Var[y_{2T+1} - \hat{y}_{2T+1}] + Var[\epsilon_{2T+2} - \hat{\epsilon}_{2T+2}] \\
&= a_{22}^2 Var[y_{2T+1} - \hat{y}_{2T+1}] + Var[\epsilon_{2T+2} - \rho_{12}\frac{\sigma_2}{\sigma_1}\hat{\epsilon}_{1T+2}] \\
&= a_{22}^2(1 - \rho_{12}^2)\sigma_2^2 + (1 - \rho_{12}^2)\sigma_2^2 \\
&= (1 + a_{22}^2)(1 - \rho_{12}^2)\sigma_2^2.
\end{aligned}
$$

At $T + 3$, equation (7) gives

$$
\epsilon_{1T+3} = y_{1T+3} - (a_{11}y_{1T+2} + a_{12}y_{2T+2}).
$$

Since we do not observe y_{2T+2}, we do not observe ϵ_{1T+3}. Again, there are two ways to make a prediction on ϵ_{1T+3}. One way is to set $E(\epsilon_{1T+3}) = 0$ and to set the variance of ϵ_{1T+3} be the first element of the MSE of the ordinary bivariate VAR forecast, that is, the first element of the matrix $\Sigma_{i=0}^{2} A^i \Omega_\epsilon A^{i'}$. The alternative way is to predict $\hat{\epsilon}_{1T+3}$ through the residual form $\hat{\epsilon}_{1T+3} = y_{1T+3} - (a_{11}y_{1T+2} + a_{12}\hat{y}_{2T+2})$. In the latter case, the variance of the difference in error becomes

$$
\begin{aligned}
Var[\epsilon_{1T+3} - \hat{\epsilon}_{1T+3}] &= a_{12}^2 Var[y_{2T+2} - \hat{y}_{2T+2}] \\
&= a_{12}^2(1 + a_{22}^2)(1 - \rho_{12}^2)\sigma_2^2.
\end{aligned}
$$

If the sufficient condition $\{a_{12}^2(1 + a_{22}^2)(1 - \rho_{12}^2)\sigma_2^2 <$ the first element of matrix $\Sigma_{i=0}^{2} A^i \Omega_\epsilon A^{i'}\}$ holds, we would use $\hat{\epsilon}_{1T+3}$ rather than $E(\epsilon_{1T+3}) = 0$. Then notify that

$$
\begin{aligned}
y_{2T+3} &= a_{21}y_{1T+2} + a_{22}y_{2T+2} + \epsilon_{2T+3} \\
\hat{y}_{2T+3} &= a_{21}y_{1T+2} + a_{22}\hat{y}_{2T+2} + \hat{\epsilon}_{2T+3}.
\end{aligned}
$$

By equation (3), the variance of the forecast error is followed by

$$
\begin{aligned}
Var[y_{2T+3} - \hat{y}_{2T+3}] &= a_{22}^2 Var[y_{2T+2} - \hat{y}_{2T+2}] + Var[\epsilon_{2T+3} - \hat{\epsilon}_{2T+3}] \\
&= a_{22}^2 Var[y_{2T+2} - \hat{y}_{2T+2}] + Var[\epsilon_{2T+3} - \rho_{12}\frac{\sigma_2}{\sigma_1}\hat{\epsilon}_{1T+3}] \\
&= a_{22}^2(1 + a_{22}^2)(1 - \rho_{12}^2)\sigma_2^2 + (1 - \rho_{12}^2)\sigma_2^2 \\
&= (1 + a_{22}^2 + a_{22}^4)(1 - \rho_{12}^2)\sigma_2^2.
\end{aligned}
$$

Iterating forward, we need to check if the sufficient condition of $\{(\Sigma_{i=0}^{k-2} a_{22}^{2i})a_{12}^2)(1 - \rho_{12}^2)\sigma_2^2 <$ the first element of matrix $\Sigma_{i=0}^{k-1} A^i \Omega_\epsilon A^{i'}\}$ holds. If this sufficient condition holds,

then the forecast MSE of the updated bivariate VAR given s more periods real-time information is followed by

$$MSE^u[y_{2T}(k)] = (\sum_{i=0}^{k-1} a_{22}^{2i})(1 - \rho_{12}^2)\sigma_2^2, \qquad 2 \leq k \leq s.$$

∎

Proof of proposition 4. The full information set is $\mathcal{I}_5$. Rearrange equation (15) at time $T + 1$, the following equation holds.

$$\epsilon_{mT+1} = y_{mT+1} - a_{m1}y_{1T} - a_{m2}y_{2T} - \dots - a_{mN}y_{NT}.$$

Given the known information set $\mathcal{I}_5$, we know ϵ_{mT+1}. All the assumptions hold, The multivariate prediction equation becomes

$$\epsilon_{jt} = \alpha_1\epsilon_{1t} + \alpha_2\epsilon_{2t} + \dots + \alpha_m\epsilon_{mt} + e_{jt} \qquad t = 1, \dots, T \quad j = m+1, \dots, N,$$

where the $\alpha_1, \dots, \alpha_m$ are the population least squares coefficients. We project the ϵ_{jt} onto **all available error information** $\epsilon_{1t}, \dots, \epsilon_{mt}$. In the multivariate normal case, vector $\epsilon_i = (\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_m)'$ and vector $\epsilon_j = (\epsilon_{m+1} \ \epsilon_{m+2} \ \dots \ \epsilon_N)'$ are considered as

$$\begin{pmatrix} \epsilon_i \\ \epsilon_j \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{\epsilon_i\epsilon_i} & \Omega_{\epsilon_i\epsilon_j} \\ \Omega_{\epsilon_i\epsilon_j} & \Omega_{\epsilon_j\epsilon_j} \end{pmatrix} \right).$$

where the partition matrix $\Omega_{\epsilon_i\epsilon_i}$ with the dimension of $m \times m$ is

$$\Omega_{\epsilon_i\epsilon_i} = \begin{pmatrix} Var(\epsilon_1) & Cov(\epsilon_1, \epsilon_2) & \dots & Cov(\epsilon_1, \epsilon_m) \\ Cov(\epsilon_2, \epsilon_1) & Var(\epsilon_2) & \dots & Cov(\epsilon_2, \epsilon_m) \\ \dots & & & \\ Cov(\epsilon_m, \epsilon_1) & Cov(\epsilon_m, \epsilon_2) & \dots & Var(\epsilon_m) \end{pmatrix},$$

the $(N - m) \times (N - m)$ partition matrix $\Omega_{\epsilon_j\epsilon_j}$ are

$$\Omega_{\epsilon_j\epsilon_j} = \begin{pmatrix} Var(\epsilon_{m+1}) & Cov(\epsilon_{m+1}, \epsilon_{m+2}) & \dots & Cov(\epsilon_{m+1}, \epsilon_N) \\ Cov(\epsilon_{m+2}, \epsilon_{m+1}) & Var(\epsilon_{m+2}) & \dots & Cov(\epsilon_{m+2}, \epsilon_N) \\ \dots & & & \\ Cov(\epsilon_N, \epsilon_{m+1}) & Cov(\epsilon_N, \epsilon_{m+2}) & \dots & Var(\epsilon_N) \end{pmatrix},$$

and $m \times (N - m)$ partition matrix $\Omega_{\epsilon_i \epsilon_j}$ is

$$\Omega_{\epsilon_i \epsilon_j} = \begin{pmatrix} Cov(\epsilon_1, \epsilon_{m+1}) & Cov(\epsilon_1, \epsilon_{m+2}) & ... & Cov(\epsilon_1, \epsilon_N) \\ Cov(\epsilon_2, \epsilon_{m+1}) & Cov(\epsilon_2, \epsilon_{m+2}) & ... & Cov(\epsilon_2, \epsilon_N) \\ & ... & & \\ Cov(\epsilon_m, \epsilon_{m+1}) & Cov(\epsilon_m, \epsilon_{m+2}) & ... & Cov(\epsilon_m, \epsilon_N) \end{pmatrix}.$$

The partition matrix of $\Omega_{\epsilon_j \epsilon_i}$ is the transpose of the partition matrix $\Omega_{\epsilon_i \epsilon_j}$. Thus, by the definition of Schur complement,[14] the conditional variance ϵ_j given ϵ_i is the Schur complement of $\Omega_{\epsilon_i \epsilon_i}$ in Ω. That is,

$$Var(\epsilon_j | \epsilon_i) = \Omega_{\epsilon_j \epsilon_j} - \Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1} \Omega'_{\epsilon_i \epsilon_j},$$

where the matrix $\Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1}$ is known as the matrix of regression coefficients. This yields least squared estimate of α,

$$\hat{\alpha} = \Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1}.$$

This is a $((N - m) \times m)$ dimensional matrix.

As a result, the general prediction form of the random column vector ϵ_j conditional on the random column vector ϵ_i is

$$\hat{\epsilon}_{jT+1} = \Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1} \epsilon_{iT+1}.$$

Then the forecast error for forecasting $\hat{y}_{jT+1}$ becomes

$$y_{jT+1} - \hat{y}_{jT+1} = \epsilon_{jT+1} - \Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1} \epsilon_{iT+1}.$$

The corresponding general form of mean squared error $MSE^u[y_{jT}(1)]$ for forecasting y_{jT+1} is

$$\begin{aligned} MSE^u[y_{jT}(1)] &= Var(y_{jT+1} - \hat{y}_{jT+1}) \\ &= Var(\epsilon_{jT+1} - \Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1} \epsilon_{iT+1}) \\ &= Var(\epsilon_{jT+1}) + Var(\Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1} \epsilon_{iT+1}) - 2Cov(\epsilon_{jT+1}, \Omega_{\epsilon_j \epsilon_i} \Omega_{\epsilon_i \epsilon_i}^{-1} \epsilon_{iT+1}). \end{aligned}$$

∎

Proof of proposition 5. This proof is done by induction. Firstly we show that the forecast MSE knowing the values of two variables is less than the forecast MSE with knowing one

---

[14]http://en.wikipedia.org/wiki/Schur_complement

value of variable, that is, $MSE[y_{4t}^{1:2}(1)] < MSE[y_{4t}^{1}(1)]$. By equation (10), the forecast MSE with knowing one value of variable is

$$MSE[y_{4t}^{1}(1)] = (1 - \rho_{14}^2)\sigma_4^2.\tag{26}$$

By equation (13), the forecast MSE with knowing the value of two variables is

$$MSE[y_{4t}^{1:2}(1)] = (1 - \underbrace{\frac{\rho_{14}^2 + \rho_{24}^2 - 2\rho_{14}\rho_{24}\rho_{12}}{1 - \rho_{12}^2}}_{E})\sigma_4^2.\tag{27}$$

To show equation (26) is larger than equation (27), we need to show that $\frac{\rho_{14}^2 + \rho_{24}^2 - 2\rho_{14}\rho_{24}\rho_{12}}{1 - \rho_{12}^2} > \rho_{14}^2$. Since

$$\frac{\rho_{14}^2 + \rho_{24}^2 - 2\rho_{14}\rho_{24}\rho_{12}}{1 - \rho_{12}^2} - \rho_{14}^2 = \frac{(\rho_{24} - \rho_{12}\rho_{14})^2}{1 - \rho_{12}^2}$$
$$\geq 0,$$

this proofs $MSE[y_{4t}^{1:2}(1)] < MSE[y_{4t}^{1}(1)]$.

Secondly, we show that the forecast MSE with knowing the values of three variables is less than the forecast MSE with knowing the values of three variables, that is, $MSE[y_{4t}^{1:2:3}(1)] < MSE[y_{4t}^{1:2}(1)]$. From equation (17),

$$MSE[y_{4t}^{1:2:3}(1)] = VAR(\epsilon_{4T+1} - \Omega_{\epsilon_4\epsilon_3}\Omega_{\epsilon_3\epsilon_3}^{-1}\epsilon_{iT+1}).\tag{28}$$

Now let's look it term by term. The (1×3) partition matrix $\Omega_{\epsilon_4\epsilon_3}$ is

$$\Omega_{\epsilon_4\epsilon_3} = \left(\begin{array}{ccc} \rho_{41}\sigma_4\sigma_1 & \rho_{42}\sigma_4\sigma_2 & \rho_{43}\sigma_4\sigma_3 \end{array}\right).$$

The (3×1) partition matrix ϵ_{iT+1} is

$$\epsilon_{iT+1} = \left(\begin{array}{c} \epsilon_{1T+1} \\ \epsilon_{2T+1} \\ \epsilon_{3T+1} \end{array}\right).$$

The (3×3) partition matrix $\Omega_{\epsilon_3\epsilon_3}$ is

$$
\Omega_{\epsilon_3\epsilon_3} = \begin{pmatrix}
Var(\epsilon_1) & Cov(\epsilon_1, \epsilon_2) & Cov(\epsilon_1, \epsilon_3) \\
Cov(\epsilon_2, \epsilon_1) & Var(\epsilon_2) & Cov(\epsilon_2, \epsilon_3) \\
Cov(\epsilon_3, \epsilon_1) & Cov(\epsilon_3, \epsilon_2) & Var(\epsilon_3)
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\
\rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\
\rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2
\end{pmatrix}.
$$

The inverse matrix of $\Omega_{\epsilon_i\epsilon_i}$ is

$$
\Omega_{\epsilon_3\epsilon_3}^{-1} = \frac{1}{(1 - \rho_{23}^2 + \rho_{12}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{13}(\rho_{12}\rho_{23} - \rho_{13}))\sigma_1^2\sigma_2^2\sigma_3^2} \times
$$

$$
\begin{pmatrix}
(1 - \rho_{23}^2)\sigma_2^2\sigma_3^2 & (\rho_{23}\rho_{31} - \rho_{12})\sigma_1\sigma_2\sigma_3^2 & (\rho_{12}\rho_{23} - \rho_{13})\sigma_1\sigma_2^2\sigma_3 \\
(\rho_{23}\rho_{31} - \rho_{12})\sigma_1\sigma_2\sigma_3^2 & (1 - \rho_{13}^2)\sigma_1^2\sigma_3^2 & (\rho_{13}\rho_{21} - \rho_{23})\sigma_1^2\sigma_2\sigma_3 \\
(\rho_{12}\rho_{23} - \rho_{13})\sigma_1\sigma_2^2\sigma_3 & (\rho_{13}\rho_{21} - \rho_{23})\sigma_1^2\sigma_2\sigma_3 & (1 - \rho_{12}^2)\sigma_1^2\sigma_2^2
\end{pmatrix}.
$$

Putting the above all partition matrix together, equation (28) becomes

$$
\begin{aligned}
MSE[y_{4t}^{1:2:3}(1)] &= VAR(\epsilon_{4T+1} - \Omega_{\epsilon_4\epsilon_3}\Omega_{\epsilon_3\epsilon_3}^{-1}\epsilon_{iT+1}) \\
&= VAR(\epsilon_{4T+1} - \underbrace{\frac{\rho_{41}(1 - \rho_{23}^2) + \rho_{42}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{43}(\rho_{12}\rho_{23} - \rho_{13})}{1 - \rho_{23}^2 + \rho_{12}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{13}(\rho_{12}\rho_{23} - \rho_{13})}}_{A} \frac{\sigma_4}{\sigma_1}\epsilon_{1T+1} \\
&\quad - \underbrace{\frac{\rho_{41}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{42}(1 - \rho_{13}^2) + \rho_{43}(\rho_{13}\rho_{21} - \rho_{23})}{1 - \rho_{23}^2 + \rho_{12}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{13}(\rho_{12}\rho_{23} - \rho_{13})}}_{B} \frac{\sigma_4}{\sigma_2}\epsilon_{2T+1} \\
&\quad - \underbrace{\frac{\rho_{41}(\rho_{12}\rho_{23} - \rho_{13}) + \rho_{42}(\rho_{13}\rho_{21} - \rho_{23}) + \rho_{43}(1 - \rho_{12}^2)}{1 - \rho_{23}^2 + \rho_{12}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{13}(\rho_{12}\rho_{23} - \rho_{13})}}_{C} \frac{\sigma_4}{\sigma_3}\epsilon_{3T+1}) \\
&= (1 + A^2 + B^2 + C^2 - 2A\rho_{41} - 2B\rho_{42} - 2C\rho_{43} + 2AB\rho_{12} + 2AC\rho_{31} + 2BC\rho_{23})\sigma_4^2 \\
&= (1 - \frac{A\rho_{41} + B\rho_{42} + C\rho_{43}}{1 - \rho_{23}^2 + \rho_{12}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{13}(\rho_{12}\rho_{23} - \rho_{13})})\sigma_4^2 \\
&= (1 - \underbrace{\frac{\rho_{41}^2(1 - \rho_{23}^2) + \rho_{42}^2(1 - \rho_{13}^2) + \rho_{43}^2(1 - \rho_{12}^2) + 2\rho_{41}\rho_{42}(\rho_{23}\rho_{31} - \rho_{21})}{1 - \rho_{23}^2 + \rho_{12}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{13}(\rho_{12}\rho_{23} - \rho_{13})}}_{D} \\
&\quad + \underbrace{\frac{2\rho_{41}\rho_{43}(\rho_{12}\rho_{23} - \rho_{13}) + 2\rho_{42}\rho_{43}(\rho_{13}\rho_{12} - \rho_{23})}{1 - \rho_{23}^2 + \rho_{12}(\rho_{23}\rho_{31} - \rho_{21}) + \rho_{13}(\rho_{12}\rho_{23} - \rho_{13})}}_{D})\sigma_4^2.
\end{aligned}
$$

Now to show $MSE[y_{4t}^{1:2:3}(1)] < MSE[y_{4t}^{1:2}(1)]$, we need to show $D > E$. Let's look at the denominator first. The common factor of the denominator is $(1 - \rho_{12}^2)(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23})$. From equation (13), we know that $\rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23}/(1 - \rho_{12}^2)$ must smaller than 1. Hence, it must be the case that $1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} > 0$. It is obvious that $1 - \rho_{12}^2 > 0$ by the stationary assumption. Therefore, the denominator is positive.

After multiplying the common factor, the nominator becomes $\{\rho_{12}^2\rho_{42}^2\rho_{13}^2 + 2\rho_{41}\rho_{42}\rho_{23}\rho_{13}(1 + \rho_{12}^2) + \rho_{41}^2\rho_{23}^2\rho_{12}^2 - 2\rho_{12}\rho_{23}\rho_{13}(\rho_{41}^2 + \rho_{42}^2) - 2\rho_{41}\rho_{42}\rho_{21}(\rho_{13}^2 + \rho_{23}^2) + \rho_{43}^2(1 - \rho_{12}^2)^2 + 2\rho_{41}\rho_{43}(\rho_{12}\rho_{23} - \rho_{13})(1 - \rho_{12}^2) + 2\rho_{42}\rho_{43}(\rho_{13}\rho_{12} - \rho_{23})(1 - \rho_{12}^2) + \rho_{41}^2\rho_{13}^2 + \rho_{42}^2\rho_{23}^2\}$. Rearranging this long expression, we have $(\rho_{12}\rho_{42}\rho_{13} + \rho_{41}\rho_{23}\rho_{12} + \rho_{43}(1 - \rho_{12}^2) - \rho_{41}\rho_{13} - \rho_{42}\rho_{23})^2$, which is nonnegative. Combining both the nominator and the denominator, we conclude that $D > E$. This proofs $MSE[y_{4t}^{1:2:3}(1)] < MSE[y_{4t}^{1:2}(1)]$.

By induction, we conclude that $MSE[y_{jt}^{1:2:...m}(1)] < MSE[y_{jt}^{1:2:...(m-1)}(1)] < ... < MSE[y_{jt}^{1:2}(1)] < MSE[y_{jt}^{1}(1)]$. ∎

Proof of proposition 6. Given $\sigma_1 = \sigma_2 = \sigma_3 = 1$, equations (18), (19), (20), (21), (22), and (23) become

$$S^{1:2} = (1 - \rho_{12}^2) + (1 - \rho_{13}^2) + (1 - \frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{12}^2})$$

$$S^{2:1} = (1 - \rho_{12}^2) + (1 - \rho_{23}^2) + (1 - \frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{12}^2})$$

$$S^{1:3} = (1 - \rho_{13}^2) + (1 - \rho_{12}^2) + (1 - \frac{\rho_{12}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{13}^2})$$

$$S^{3:1} = (1 - \rho_{13}^2) + (1 - \rho_{23}^2) + (1 - \frac{\rho_{12}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{13}^2})$$

$$S^{2:3} = (1 - \rho_{23}^2) + (1 - \rho_{12}^2) + (1 - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2})$$

$$S^{3:2} = (1 - \rho_{23}^2) + (1 - \rho_{13}^2) + (1 - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2}).$$

Considering $|\rho_{13}| > |\rho_{12}| > |\rho_{23}|$, we see directly from above equations that $min\{S^{1:2}, S^{2:1}\} =$

$S^{1:2}$, $min\{S^{1:3}, S^{3:1}\} = S^{1:3}$, and $min\{S^{2:3}, S^{3:2}\} = S^{3:2}$.

Comparing $S^{1:2}$ to $S^{1:3}$, it is the case that $S^{1:2}$ is less than $S^{1:3}$ if we can show that

$$\frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{12}^2} > \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2}.$$

Now

$$\frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{12}^2} - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2} = \frac{(\rho_{13}^2 - \rho_{12}^2)(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23})}{(1 - \rho_{12}^2)(1 - \rho_{23}^2)}$$
$$> 0,$$

since $|\rho_{13}| > |\rho_{12}|$ by assumption and $(\rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23})/(1 - \rho_{12}^2) < 1$ by equation (13). This proofs $min\{S^{1:2}, S^{1:3}\} = S^{1:2}$. By symmetry, $min\{S^{1:2}, S^{3:2}\} = S^{1:2}$. Therefore, $min\{S^{1:2}, S^{2:1}, S^{1:3}, S^{3:1}, S^{2:3}, S^{3:2}\} = S^{1:2}$. ∎

Proof of proposition 7. The purpose of this proposition is to find minimum S, that is, $min\{S^{1:2}, S^{2:1}, S^{1:3}, S^{3:1}, S^{2:3}, S^{3:2}\}$. Firstly, we compare equation (18) to equation (19). Since $\sigma_2 > \sigma_1$, the term one of equation (18) is larger than that of equation (19). Because $\rho_{23} > \rho_{13}$, the term two of equation (18) is bigger than that of equation (19). The third term of both equation is the same. It end up with $min\{S^{1:2}, S^{2:1}\} = S^{2:1}$. By the same method and symmetry in the equations (20), (21) and (22), (23), we find that $min\{S^{1:3}, S^{3:1}\} = S^{3:1}$ and $min\{S^{2:3}, S^{3:2}\} = S^{2:3}$.

Secondly, we want to find $min\{S^{2:1}, S^{3:1}, S^{2:3}\}$. To see $min\{S^{2:1}, S^{3:1}\}$, the first one terms of expression (19) is smaller than that of expression (21) since $\rho_{12} > \rho_{13}$ by assumption. Now we combine the last two terms for expression (19), we have

$$\frac{2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{13} + \rho_{12}\rho_{23})^2}{1 - \rho_{12}^2}\sigma_3^2 \tag{29}$$

Likewise, we combine the last two terms for expression(21), we have

$$\frac{2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{12} + \rho_{13}\rho_{23})^2}{1 - \rho_{13}^2}\sigma_2^2 \tag{30}$$

If we can show expression (29) is smaller than expression (30), then we can find $min\{S^{2:1}, S^{3:1}\}$. Since

$$\frac{(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{13} + \rho_{12}\rho_{23})^2)(1 - \rho_{13}^3)}{(1 - \rho_{12}^2)(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{12} + \rho_{13}\rho_{23})^2)}\frac{\sigma_3^2}{\sigma_2^2} <$$
$$\frac{(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{13} + \rho_{12}\rho_{23})^2)(1 - \rho_{13}^3)}{(1 - \rho_{12}^2)(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{12} + \rho_{13}\rho_{23})^2)}$$

Now since

$$
\begin{aligned}
&(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{13} + \rho_{12}\rho_{23})^2)(1 - \rho_{13}^3) - \\
&(1 - \rho_{12}^2)(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{12} + \rho_{13}\rho_{23})^2) \\
&= \rho_{12}^2(1 - \rho_{12}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}) - \rho_{13}^2(1 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}) \\
&< \rho_{12}^2 - \rho_{12}^2(1 - \rho_{13}^2) - \rho_{13}^2 + \rho_{13}^2(1 - \rho_{12}^2) \\
&= 0
\end{aligned}
$$

So $(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{13} + \rho_{12}\rho_{23})^2)(1 - \rho_{13}^3) < (1 - \rho_{12}^2)(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{12} + \rho_{13}\rho_{23})^2)$, which gives

$$
\frac{(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{13} + \rho_{12}\rho_{23})^2)(1 - \rho_{13}^3)}{(1 - \rho_{12}^2)(2(1 - \rho_{13}^2 - \rho_{23}^2 - \rho_{12}^2) + (\rho_{12} + \rho_{13}\rho_{23})^2)} < 1.
$$

This proofs expression (29) is smaller than expression (30). Therefore, $min\{S^{2:1}, S^{3:1}\} = S^{2:1}$.

Finally, we need to find $min\{S^{2:1}, S^{2:3}\}$. Since the first two terms of equation (19) and equation (22) are the same, we only need to compare the last term of two equations.

$$
1 - \frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{12}^2}\sigma_3^2 = \frac{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{12}^2}\sigma_3^2
$$

and

$$
1 - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{23}^2}\sigma_1^2 = \frac{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{13}^2}\sigma_1^2.
$$

Because $\sigma_3^2 > \sigma_1^2$ and $\rho_{12} > \rho_{23}$,

$$
\frac{\sigma_3^2}{1 - \rho_{12}^2} > \frac{\sigma_1^2}{1 - \rho_{12}^2} > \frac{\sigma_1^2}{1 - \rho_{23}^2}.
$$

This proofs $min\{S^{2:1}, S^{2:3}\} = S^{2:3}$.

Therefore, the firm who has the largest variance and relatively bigger correlation to least variance firm releases first. In other words, suppose $\sigma_2 > \sigma_3 > \sigma_1$, $\rho_{21} > \rho_{23}$, and $\rho_{21} > \rho_{31}$, then the minimum of the sum of squared forecast errors is $min\{S^{1:2}, S^{2:1}, S^{1:3}, S^{3:1}, S^{2:3}, S^{3:2}\} = S^{2:3}$. Thus, the informational ordering is Firm 2, followed by Firm 3 and finally Firm 1. ∎

References

Abarbanell J. S. and V. L. Bernard (1992). Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior. *Journal of Finance* 47, 1181-1207.

Beckers, S., M. E. Steliaros and A. Thomson (2004). Bias in European Analysts' Earnings Forecasts. *Financial Analysts Journal* 60, 74-85.

Begley, J. and P. E. Fischer (1998). Is There Information in an Earnings Announcement Delay? *Review of Accounting Studies* 3, 347-363.

Bown, L. D., R. L. Hagerman, and M. E. Zmijewski (1987). Security Analyst Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings. *Journal of Accounting and Economics* 9, 61-87.

Brown, L. and M. Rozeff (1979). Univariate Time-Series Models of Quarterly Accounting Earnings Per Share: a proposed premier model. *Journal of Accounting Research* Spring, 178-189.

—— and —— (1978). The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings. *Journal of Finance* March, 1-16.

Brown, Philip, Foster, George and Noreen, Eric (1985). Security Analyst Multi-Year Earnings Forecasts and the Capital Market. Sarasota: American Accounting Association.

Clement, M. B. and S. Y. TSE (2005). Financial Analyst Characteristics and Herding Behavior in Forecasting. *Journal of Finance* 60, 307-341.

Conroy, R. and R. Harris (1987). Consensus Forecasts of Corporate Earnings: Analysts' Forecasts and Time Series Methods. *Management Science* 33:6, 725-738.

Cragg, J. G. and B. G. Malkiel (1968). The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings. *Journal of Finance* 23:1 67-84.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods.* New York, Oxford University Press.

De Bondt, W. F. M., and R. H. Thaler (1985). Does the Stock Market Overreact? *Journal of Finance* 40, 793-805.

—— and —— (1990). Do Security Analysts Overreact? *A.E.R. Papers and Proc* 80, 52-57.

Dogan, M., E. Coskun, and O. Celik (2007). Is Timing of Financial Reporting Related to Firm Performance? An Examination on Ise Listed Companies. *International Research Journal of Finance and Economics* 12, 220-233.

Doucet, A., de Freitas N, and Gordon NJ (2001). *Sequential Monte Carlo Methods in Practice.* Springer: New York.

Easterwood, J. C. and S. R. Nutt (1999). Inefficiency in Analysts' Earnings Forecasts: Systematic Misreaction or Systematic Optimism? *Journal of Finance* 54, 1777-1797.

Elton, E. J. and M. J. Gruber (1972). Earnings Estimates and the Accuracy of Expectational Data. *Management Science* 18, 409-424.

Ertimur, Y., J Sunder, and S. V. Sunder (2007). Measure for Measure: The Relation between Forecast Accuracy and Recommendation Profitability of Analysts. *Journal of Accounting Research* 45, 567-606.

Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The Economic Implications of Corporate Financial Reporting. *Journal of Accounting and Economics* 40, 3-73.

Gregory, A. W. and H. Zhu (2008). Forecast Accuracy Improvement: Evidence from Nonfarm Payroll Employment. Working paper.

Hamilton, J.D. (1994). *Time series analysis.* Princeton University Press, Princeton.

Hong, H. and J. D. Kubik (2003). Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts. *Journal of Finance* 58, 313-351.

Hopwood W., J. McKeown and P. Newbold (1982). The Additional Information Content of Quarterly Earnings Reports: Intertemporal Disaggregation. *Journal of Accounting Research* Autumn 343-349.

Imhoff, E. A., JR., and P. V. Pare (1982). Analysis and Comparison of Earnings Forecast Agents. *Journal of Accounting Research* 20:2, 429-439.

Jarque, C. M. and A. K. Bera (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review* 55, 163-172.

Keane M. P. and D. E. Runkle (1998). Are Financial Analysts' Forecasts of Corporate Profits Rational? *Journal of Political Economy* 106:4, 768-805.

Lang, M. H. and R. J. Lundholm (1996). Corporate Disclosure Policy and Analyst Behavior. *The Accounting Review* 71:4, 467-492.

Lim, T. (2001). Rationality and Analysts' Forecast Bias. *Journal of Finance* 56, 369-385.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis.* Springer: Berlin.

—— and —— (1996). *Handbook of Matrices* John Wiley & Sons: England.

Maingot, M. and D. Zeghal (2008). An Analysis of Corporate Governance Information Disclosure by Canadian Banks. *Corporate Ownership & Control* 5:2, 225-236.

Nyblom, J. and A. Harvey (2000). Tests of Common Stochastic Trends. *Econometric Theory* 16, 176-199.

Pesaran, M. H. and M. R. Wickens (1995). *Handbook of Applied Econometrics: Macroeconomics.* Blackwell: Oxford UK & Cambridge USA.

Phillips, P.C.B. and P. Perron (1988). Testing for a Unit Root in Time Series Regressions. *Biometrika* 75, 335-346.

Sengupta, P. (2004). Disclosure Timing: Determinants of Quarterly Earnings Release Dates. *Journal of Accounting and Public Policy* 23, 457-482.

Shumway, RH. and DS. Stoffer (2000). *Time Series Analysis and its Applications.* Springer: New York.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica* 48:1, 1-48.

Stock, J. H. and M. W. Watson (2001). Vector Autoregressions. *The Journal of Economic Perspectives* 15, 101-115.

Trueman, B. (1990). Theories of Earnings Announcement Timing. *Journal of Accounting and Economics* 13:3, 285-301.

Tsay, R. S. (2005). *Analysis of Financial Time Series.* John Wiley & Sons, New Jersey.

Zhang, F. Zh. (2005). *The Schur Complement and Its Applications.* Springer.

Zhu, H. (2008). Earnings Forecasts and Timing of Announcement. Working paper.