Three Essays on Updating Forecasts in Vector Autoregression Models

by

Hui Zhu

A thesis submitted to the
Department of Economics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
April 2010

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 market data release for a specific time period of observation is often spread over several days or weeks. This thesis 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 U.S. nonfarm payroll employment and the six Canadian banks shows its value. By using the updated VAR forecast, we conclude that there are relative efficiency gains in the one-step-ahead forecast compared to the ordinary VAR forecast, and compared to professional consensus forecasts. Thought experiments emphasize that the release ordering is crucial in determining forecast accuracy.

Co-Authorship

All publications arising from this work will be joint with my supervisor, Allan W. Gregory, at the Department of Economics, Queen's University

Dedication

To my parents

for letting me pursue my dream

for so long so far away from home

And

To my son

for giving me new dreams to pursue

Acknowledgments

I am deeply indebted to my supervisor Professor Allan W. Gregory whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis. I sincerely acknowledge his generous time to ensure the quality of this thesis and the timely completion. I am also grateful to Professors Christopher Ferrall, Allen Head, Susumu Imai, James MacKinnon, Morten Nielsen, Katsumi Shimotsu, and Gregor Smith, for much helpful advice. I would like to thank Professors Weili Ding and Steven Lehrer from Queen's School of Policy Studies who have long been an inspiration to me.

I would like to express my gratitude to my many friends and colleagues at Queen's with whom I have had the pleasure of working over the years. These include David Byrne, Pamela Cardozo, Adnan Khan, Afrasiab Mirza, Lealand Morin, Salem Nechi, Eric Stephens, Ryan Webb, and Tommy Wu.

I would like to thank all those who gave me the possibility to complete this thesis. I would also like to thank my parents, my sister and brother, and other family members to pursue my studies. Especially, I would like to give my special thanks to my son Steven whose patient love enabled me to complete this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Individual analysts' forecasting of market data is available widely on the internet, such as Bloomberg and *The Wall Street Journal*. In general, economic calendars stay up-to-date on key releases with consensus forecasts. In addition, firms' earnings calendars sorted by different industries are available online. A snapshot of Bloomberg's online screen on June 2, 2009 is shown as follows. Analysts announce monthly or quarterly outcomes (firms' earnings, sales, GDP, inflation, and employment, *etc.*) on a specific date every month or quarter. The dates are predetermined at the beginning of the year for each month or quarter. One or two weeks before each announcement date, individual professional forecasters often make a forecast for each firm or macroeconomic indicator, which are typically proprietary. Simultaneously a consensus forecast, averaging individual forecasts, usually appears online one or two weeks in advance of this date. Based on the specific announcement dates, the market participants and investors achieve the most up-to-date financial information about the firm or economy.

Bloomberg.com

ECONODAY

Market Focus

Tuesday: Another gain in pending home sales would increase talk that the sector has finally bottomed out.

Monthly | Weekly | Daily | Today | 4:54 PM ET

Jan Feb Mar Apr May **Jun** Jul Aug Sep Oct Nov Dec

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Monday Jun 1	**Tuesday Jun 2**	Wednesday Jun 3	Thursday Jun 4	Friday Jun 5
Market Focus » **Market Reflections »** *2-Yr Note Settlement* *5-Yr Note Settlement* *7-Yr Note Settlement* Personal Income and Outlays REPORT 8:30 AM ET ISM Mfg Index REPORT 10:00 AM ET Construction Spending REPORT 10:00 AM ET 4-Week Bill Announcement REPORT 11:00 AM ET 3-Month Bill Auction REPORT 1:00 PM ET 6-Month Bill Auction REPORT 1:00 PM ET **Timonthy Geithner Speaks** 9:15 PM ET	**Market Focus »** **Market Reflections »** Motor Vehicle Sales REPORT ICSC-Goldman Store Sales REPORT 7:45 AM ET Redbook REPORT 8:55 AM ET Pending Home Sales Index REPORT 10:00 AM ET 4-Week Bill Auction REPORT 1:00 PM ET 52-Week Bill Auction REPORT 1:00 PM ET	**Market Focus »** Bank Reserve Settlement MBA Purchase Applications 7:00 AM ET Challenger Job-Cut Report 7:30 AM ET ADP Employment Report 8:15 AM ET **Ben Bernanke Speaks** 10:00 AM ET Factory Orders CONSENSUS 10:00 AM ET ISM Non-Mfg Index CONSENSUS 10:00 AM ET EIA Petroleum Status Report 10:30 AM ET **Thomas Hoenig Speaks** 2:30 PM ET	*Weekly Bill Settlement* *52-Week Bill Settlement* Chain Store Sales Monster Employment Index **Sandra Pianalto Speaks** 7:50 AM ET **William Dudley Speaks** 8:00 AM ET Jobless Claims CONSENSUS 8:30 AM ET Productivity and Costs CONSENSUS 8:30 AM ET **Ben Bernanke Speaks** 8:45 AM ET 30-Yr Bond Announcement 9:00 AM ET EIA Natural Gas Report 10:30 AM ET 3-Month Bill Announcement 11:00 AM ET 6-Month Bill Announcement 11:00 AM ET 3-Yr Note Announcement 11:00 AM ET 10-Yr Note Announcement 11:00 AM ET Treasury STRIPS 3:00 PM ET Fed Balance Sheet 4:30 PM ET Money Supply 4:30 PM ET	Employment Situation CONSENSUS 8:30 AM ET **Donald Kohn Speaks** 2:15 PM ET Consumer Credit CONSENSUS 3:00 PM ET
Equity Settlement 6-4-09	Equity Settlement 6-5-09	Equity Settlement 6-8-09	Equity Settlement 6-9-09	Equity Settlement 6-10-09

ECONODAY

Actual Data Source: Haver Analytics | Consensus Data Sources: Market News International & Thomson Financial

Econoday Suggestion Box: We welcome your ideas on how we can serve you better.

Market Moving Indicator
Merit Extra Attention
Other Key Indicator
CONSENSUS Consensus Info Available
REPORT Final Data and Analysis Available

Individual forecasters in real-time confront the same prediction problems we address in this thesis. The timing of market 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 thesis 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 yet to be addressed in the literature.

Traditional forecasting in a multivariate time series setting is usually studied in vector autoregression (VAR) models. In this set-up the VAR is specified and estimated with a common end-point. 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 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 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 scenario 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 or superior 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 credibly be 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 magnitude of the earnings of a related firm on a subsequent announcement 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 major application in this thesis (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. Also such obvious manipulation of announced earnings would come under scrutiny by both regulators and investors.

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. Interestingly, for the most part, forecasting in finance has relied on univariate methods focussing on conditional volatility.

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, for a given time period most available macroeconomic variables or financial time series 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, the employment, unemployment rate, interest rates, and so on, released for the same month, are also available to the public on 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. Before the advent of the internet and the ability to access earnings announcements in real-time, it would have been almost impossible to do such an exercise.

This thesis is organized in the following manner. In Chapter 2, we study forecasting with sequential release of information in a VAR framework under various scenarios. 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) of the forecast. We first consider the simplest 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 the previous period represents the "new" information in the release information.

This difference provides an estimate of the current innovation of the release. If the innovation is correlated with the innovation of the remaining variables, then there can be a meaningful update. When the absolute correlation coefficient of the innovations between two variables is high, the updated one-period-ahead bivariate VAR forecast has a smaller MSE than the ordinary bivariate VAR forecast. In extreme cases when the correlation coefficient approaches zero, then there is no improvement (change) in the forecast and the MSE is identical to that of the standard VAR. Of course, for multistep-ahead forecasts, having the additional information will always lead to improved forecasts with lower MSEs.

The second scenario we consider is updating 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. Interestingly, there are instances when forecasts using the new information can actually lead to poorer predicted outcomes. The intuition is that the (conditional) estimate of the second period's innovation for the release variable depends on the accuracy of the first period variables in the release equation. That the variance of the second period estimates of the release innovation exceeds the unconditional variance (standard VAR) is entirely possible.

Finally, we conduct the extension to the multivariate VAR forecast with sequential release dates. With only one available real-time variable known in advance, the method 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 one-period-ahead forecast MSE for a variable monotonically decreases as other related variables are sequentially released.

In Chapter 3 of the thesis, we investigate how to use the simplest theoretical framework addressed in Chapter 2 to forecast U.S. nonfarm payroll employment from the Bureau of Labor Statistics (BLS) using the earlier (two days) release of the Automatic Data Processing Inc. (ADP). Although there are various macroeconomic announcement series or earnings announcements for firms we could consider as an application of the research, we intentionally chose a macroeconomic real variable, nonfarm payroll employment in the United States, as the first example of this method. One reason for making this choice is that total nonfarm payroll employment is the first major economic indicator released each month and is anxiously awaited by financial market participants. This is an economic report that can move the markets. Another reason is that the employment situation has frequently been used not only in the formulation of Federal Reserve policy but also as an explanation of anomalous stock price behavior. In addition, employment announcements have tractable and fixed announcement timing with an order of informational release that is invariant.

We have almost nine years of monthly nonfarm payroll data dating from January 2001 through May 2009. One important common issue to all forecasting exercises is handling of macroeconomic variables which are subject to data revisions in subsequent periods. To accommodate data revisions, we follow the conventional approach of using the sample data available at the time period the forecast is made. That is, our sample indicates the first released data, the second released data, the third released data, and the final released data, in turn. Rather than use final released data to estimate and forecast, we employ data from as many different releases as there are dates in the sample. More specifically, at every date within a sample, both right-side and left-side variables in the bivariate VAR model are the most up-to-date estimates of variables at that time. This implies that each sequential forecasting exercise has its own set of data representing the most current information available at the time this forecast occurred.

In the labor forecasting application in Chapter 3, the updated VAR has tighter confidence intervals for BLS one- and two-step-ahead forecasts than the ordinary VAR. The actual value lies in or outside the 99% asymptotic confidence intervals, but lies all in the 99% bootstrap confidence intervals. Comparing the asymptotic confidence intervals to the bootstrap confidence intervals, we find that the bootstrap has tighter confidence intervals for BLS forecasts. The labor forecasting application results also indicate that the root mean squared error in the updated VAR is less than in the ordinary VAR. The updated VAR improves 8% in a one-step-ahead forecast and 16% in a two-step-ahead forecast respectively, compared to the ordinary VAR. The results from both asymptotic and bootstrap confidence intervals of one- and two-step-ahead forecasts indicate that confidence interval for the ordinary VAR forecast is tighter than that of the updated VAR forecasting. The Monte Carlo simulation experiment demonstrates 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, it follows, is the average standard deviation. In particular when the correlation coefficient approaches one, we have perfect linear association, and the updated bivariate VAR forecast is extremely close to the actual value. Overall the evidence suggests that the asymptotic and bootstrap confidence intervals are similar. Our guess is that for more complicated VARs with additional lags, the bootstrap procedure would likely dominate.

The forecasts we do in Chapter 3 are not the only ones that face such revision issues. Many economic time series are subject to revision. Revisions to measures of real economic activity may occur immediately in the next month or years after official figures are first released. Aruoba (2008) documents the empirical properties of revisions to major macroeconomic variables in the United States. He finds that their revisions do not have a zero mean, which indicates that the initial announcements by Statistical Agencies are biased. Croushore and Stark (2001, 2003) show how data revisions can affect forecasting. They use a real-time data set to analyze data revisions. While the results of some studies are conflicting, Koenig *et al* (2003) show first-release data are to be preferred for estimation even if the analyst is ultimately interested in predicting revised data. They provide three alternative strategies for estimating forecasting equations with real-time data. They conclude that using first released data in both sides of the equation provides superior forecasting to that obtained from final released data. This thesis employs the most up-to-date estimates available at forecasting time in addition to using the updated VAR forecast method.

Earnings, on the other hand, typically do not have revisions from the first release. Our second main application, explored in Chapter 4, involves forecasting the earnings for the six large Canadian Banks. One use of earnings or macroeconomic forecasts is to provide a proxy for the market expectation of a future realization. Recent work suggests that the stock market reacts to earnings or macroeconomic announcements. Some researchers study how markets respond to labor data, such as Krueger (1996) and Boyd *et al* (2005). The latter examine how stocks respond to unemployment news, which is measured as the surprise component (a residual in a regression equation). Adopting forecasts of the change in the unemployment rate to obtain the surprise component in the announcement of the unemployment rate, they find that an announcement of rising unemployment is good news for stocks during economic expansions and bad news during economic contractions. Other researchers such as Flannery *et al* (2002), Pesaran and Wickens (1995), and Rapach *et al* (2005) focus on how markets respond to other macro variables. Faust *et al* (2007) study how U.S. macroeconomic announcements affect joint movements of exchange rates and interest rates. Using a 14-year span of high-frequency data, they conclude that unexpectedly strong announcements lead either to a fall in the risk premium required for holding foreign assets or an expected net depreciation over the ensuing decade, or both. It is the importance of earnings and their sequential release in a macroeconomy that suggests this as a very appropriate application.

Chapter 4 of the thesis forecasts earnings of the Canadian banking industry using sequential release of the banks' earnings. 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 earnings per share (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 method-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).

To illustrate the general theoretical framework, we choose quarterly EPS[1] of the big six Canadian banks. There are GAAP and non-GAAP financial measures of EPS by the U.S. Securities and Exchange Commission. 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. To be consistent with banks' own management and analysis, we use non-GAAP diluted EPS in this chapter. EPS also standardizes relative magnitudes of the six 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

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

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 compute the consensus earnings forecast which is available for all these banks over some of the sample period 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 of announcements varies, and the order of announcements is different at each quarter. We also explore the ordering of release dates in a limited sense using dummy variables but maintain the assumption the ordering is exogenously determined. Similar to the payroll forecasts in the U.S., we find that the root mean squared error in the updated VAR is less than in the ordinary VAR. The updated VAR improves 33% in a one-step-ahead forecast compared to the ordinary VAR, and improves 7% in comparison to consensus forecasts.

In this study, the order of release of information is exogenous by assumption. Further if we assume also that exogenous change in order does not change the level of earnings via intercept changes, then we can consider counterfactual orderings. For instance, we can construct any alternative orderings we wish and rank orderings according to some selected criterion. A natural criterion is an ordering that maximize disclosure to market participants. In this case, the objective function is to choose the ordering 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 depends on the firms' variance and their correlation coefficients of innovations. As a result, we investigate how slight changes in the timing of earnings announcements is helpful in accurately forecasting firms' earnings. Under this ordering, a substantial decrease in MSE can be obtained in comparison to the actual ordering.

Nonfarm payroll employment and earnings forecast are but one example of the kind of application that can be considered in this time separation of markets on a daily basis. The conclusions are summarized in Chapter 5, which also sheds light on future extensions.

To summarize the organization of the thesis, Chapter 2 presents the econometric method. Chapter 3 applies the bivariate VAR model for forecasting monthly U.S. nonfarm payroll employment. A general application of the multivariate updated VAR forecasts to the Canadian banking industry and thought experiments are illustrated in Chapter 4. Simulation studies are analyzed in both Chapter 3 and Chapter 4. Chapter 5 concludes and illustrates future extensions. Appendix A provides the proofs of propositions in Chapter 2.

Chapter 2

Theoretical Framework

2.1 Introduction

In this chapter we develop a notation and establish the central theoretical results of the thesis. Consider the N multivariate stationary vector autoregression model of order 1, or VAR (1)

$$Y_t = AY_{t-1} + \epsilon_t, \qquad t = 1, 2, ..., T \tag{2.1}$$

where $Y_t = (y_{1t}, ..., y_{Nt})'$ is a $(N \times 1)$ random vector, and A is a fixed $(N \times N)$ coefficient matrix. The VAR (1) could be stacked to allow more dependence but nothing important would change. The first subscript i represents the variable ($i = 1, 2, ..., N$) and the second subscript t represents the time period which in a finite sample runs form $t = 1, 2, ..., T$. Moreover, $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{Nt})$' is a N-dimensional white noise or innovation process, that is, for all t, $E(\epsilon_t) = 0$ and $E(\epsilon_t \epsilon_t') = \Omega_\epsilon$, with the contemporaneous covariance $Cov(\epsilon_{it},\ \epsilon_{jt}) = \rho_{ij}\sigma_i\sigma_j$ for $i, j = 1, ..., N$. The covariance matrix Ω_ϵ is assumed to be nonsingular if not otherwise stated. Since we assume a stationary vector autoregressive process, the condition on each correlation coefficient $|\rho_{ij}| < 1$ must hold. Finally, σ_i is the standard deviation of innovation ϵ_i. Given the multivariate VAR model (1), the ordinary multivariate VAR one-step-ahead forecast at time 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).

Operationally, the multivariate VAR forecast is based on the estimation in 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 generalized least squares estimator since there are no cross-equation restrictions and each equation contains the same regressor. Therefore, under the usual assumptions (normality, infinite past sample, and known coefficients) forecasts based on conditional means of the VAR are optimal in the sense they have minimum MSE compared to any other forecasts. The optimality is predicated on the assumption that other forecasts use the same information as the VAR. However, the fact is that most macroeconomic or financial time series we study in multivariate VAR forecasts do not end at the same time, that is, one variable is generally available for public use a couple of days or more 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, the unemployment rate, interest rates, 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 does not distinguish release date differences assuming a common "end" period for conditioning the forecasts for all variables.

The focus of this chapter is to examine the forecasting problem when the release of information for variables comprising the VAR occur sequentially.

2.1.1 Updating Bivariate VAR Forecasts

We start with the simplest bivariate VAR with one real-time variable available in a prior to another. To simplify the discussion, we first consider the bivariate VAR forecast, where $N = 2$. Suppose y_{1t} is observable for $t = 1, \ldots T+1$ and y_{2t} is observable only up to time T. We do not need to specify the amount of "real" time that the first release is separated from the second. Once variable two is available for $T+1$, we return to the usual forecasting problem. Then the bivariate VAR (1) model is as follows:

$$\begin{aligned} y_{1t} &= a_{11} y_{1t-1} + a_{12} y_{2t-1} + \epsilon_{1t} \qquad t = 1, ..., T, T+1 \\ y_{2t} &= a_{21} y_{1t-1} + a_{22} y_{2t-1} + \epsilon_{2t} \qquad t = 1, ..., T. \end{aligned} \tag{2.2}$$

Following the conventional multivariate VAR forecasting method based on T observations (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 the 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 conventional multivariate VAR forecast MSE.

With one more piece of real-time information (y_{1T+1}) available at time period $T+1$, we observe ϵ_{1T+1} (making the usual assumption that coefficients are known). This is because $\epsilon_{1T+1} = y_{1T+1} - (a_{11}y_{1T} + a_{12}y_{2T})$. If we regress ϵ_2 on ϵ_1 conditional on observing ϵ_{1T+1}, we have $\epsilon_{2T+1|T} = E(\epsilon_{2T+1}|\epsilon_{1T+1})$. Thus, we can forecast the residual $\epsilon_{2T+1|T}$ which given our assumptions can be expressed as $\epsilon_{2T+1|T} = (\rho_{12}\sigma_2/\sigma_1)\epsilon_{1T+1}$. Hence, the one-step-ahead forecast MSE is obtained as the following proposition:

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

$$MSE^u[y_{2T}(1)] = (1-\rho_{12}^2)\sigma_2^2. \tag{2.3}$$

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 the correlation between ϵ_1 and ϵ_2, the condition $|\rho_{12}| < 1$ must hold by stationarity. We can see directly the relationship between the conventional VAR forecasting and the updating that was y_{1T+1}. The usual bivariate VAR forecast MSE gives $MSE = \sigma_2^2$. Therefore, the one-step-ahead updated bivariate VAR forecast has a smaller or equal MSE. In the case of $\rho_{12} = 0$, the MSEs are the same and there is no difference in one period ahead MSEs.

Additionally, the higher the correlation of the innovations, the smaller the MSE in the updated bivariate VAR forecast. When the correlation coefficient of the innovations between two variables is lower, the informational advantage is lessened. Notice that the magnitude of the variance in σ_1^2 plays no role in determining the MSE for forecasting y_{2T+1}. However, this result is due to looking at only the one period ahead ($k = 1$) forecasts. We can examine the $k \geq 2$ long-horizon forecast with one more piece of real-time information known in advance.

Proposition 2 *Given the information set $\{y_{11}, y_{12}, ..., y_{1T+1}, y_{21}, ..., y_{2T}\}$, the k-step-ahead forecast mean squared error matrix in the updated bivariate VAR is*

$$\begin{aligned} MSE^u[Y_T(k)] &= \sum_{i=0}^{k-2} A^i \Omega_\epsilon A^{i'} + A^{k-1} \begin{pmatrix} 0 & 0 \\ 0 & (1-\rho_{12}^2)\sigma_2^2 \end{pmatrix} A^{k-1'} \qquad (2.4) \\ &= \Omega_\epsilon + A \quad MSE^u[Y_T(k-1)] \quad A', \qquad k \geq 2, \end{aligned}$$

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 (2.4) is smaller than the recursive, ordinary bivariate VAR forecast MSE matrix:

$$\begin{aligned} MSE[Y_T(k)] &= \sum_{i=0}^{k-2} A^i \Omega_\epsilon (A^i)' + A^{k-1} \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} A^{k-1'} \\ &= MSE[Y_T(k-1)] + A^{k-1}\Omega_\epsilon A^{k-1'} \qquad k \geq 2. \end{aligned} \tag{2.5}$$

To see the difference between the updated bivariate VAR forecast MSE^u and the ordinary bivariate VAR forecast MSE, we compare equations (2.4) and (2.5):

$$MSE[Y_T(k)] - MSE^u[Y_T(k)] = A^{k-1} \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \rho_{12}^2\sigma_2^2 \end{pmatrix} A^{k-1'}. \tag{2.6}$$

In a one-step-ahead forecast ($k = 1$), equation (2.6) indicates that the gain of employing one available item of real-time information y_{1T+1} to forecast a related variable y_{2T+1} depends only 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 conventional bivariate VAR forecast for $k = 1$.

For the two-step-ahead forecast ($k = 2$) equation (2.6) becomes

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

$$= \begin{pmatrix} (a_{11}\sigma_1 + a_{12}\rho_{12}\sigma_2)^2 & (a_{11}\sigma_1 + a_{12}\rho_{12}\sigma_2)(a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2) \\ (a_{11}\sigma_1 + a_{12}\rho_{12}\sigma_2)(a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2) & (a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2)^2 \end{pmatrix}.$$

The difference between the forecast MSE of y_{2T+1} of the ordinary bivariate VAR and that of the updated bivariate VAR is $(a_{21}\sigma_1 + a_{22}\rho_{12}\sigma_2)^2$, which is greater than 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. The two sources of the improved accuracy come from: (i) using the actual value of y_{1T+1} in the second equation (as seen from $a_{21} \neq 0$) and (ii) using an improved forecast of y_{2T+1} which depends on a_{22}, ρ_{12}, and the size of the variance of the innovation ϵ_{2t}.

Furthermore, $MSE[Y_T(k)] - MSE^u[Y_T(k)]$ converges to zero as $k \to \infty$. In other words, the MSE by the updated VAR forecast converges to the MSE of the conventional VAR forecast as the forecasting horizon gets larger. Under the assumption that our VAR (1) process is stationary, the polynomial $det(I_m - Az)$[1] has no roots in or on the complex unit circle. That is equivalent to saying that all eigenvalues of parameter matrix A have modulus less than 1. By the properties of matrices[2], A^{k-1} converges to 0 as $k \to \infty$. The realistic implication is that the informational advantage of knowing y_{1T+1} dissipates over time (as a consequence of the stationarity assumption).

2.1.2 Updating Bivariate VAR Forecasts with Data from Two More Periods Known in Advance

While the usual situation is that information is released sequentially, there are instances when there may be multiple periods of time before the release of the other variables. 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

[1]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.

[2]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.

Ontario. Users are encouraged to call with any general health questions with confidential advice being given regarding any health concerns. The 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, because NACRS data are 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 full 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 that a flu outbreak or vaccine failure has occurred and a high demand for NACRS data is expected. Understanding the impact of these limitations is crucial to studying the provincial Telehealth data and their usefulness to public health and emergency services[3]. Occasionally, there are situations in which earnings are delayed for more than one period as well.

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 follows

$$\begin{aligned} y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + \epsilon_{1t} \qquad t = 1, ..., T, T+1, ..., T+s \\ y_{2t} &= a_{21}y_{1t-1} + a_{22}y_{2t-1} + \epsilon_{2t} \qquad t = 1, ..., T. \end{aligned} \tag{2.7}$$

We assume that all the assumptions of section 2.1 hold.

[3]Adam van Dijk, Don McGuinness, Elizabeth Rolland, Kieran M Moore. 2007. “Can Telehealth respiratory call volume be used as a proxy for emergency department respiratory visit surveillance by public health?”

A one-step-ahead forecast MSE in this model set up is consistent with Proposition 1. 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 (2.7), 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 to ignore the release information and to use the unconditional prediction $E(\epsilon_{1T+2}) = 0$ with variance ϵ_{1T+2} (σ_1^2). In this case the element of the MSE of the conventional bivariate VAR forecast $(\Omega_\epsilon + A\Omega_\epsilon A')$ is relevant. 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 of σ_1^2. In the former case, the variance of the difference in error becomes

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

A sufficient condition for a lower MSE is $\{a_{12}^2(1 - \rho_{12}^2)\sigma_2^2 <$ the first row and column element of matrix $\Omega_\epsilon + A\Omega_\epsilon A'\}$ and 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. The intuition here is that we need a good forecast of y_{2T+1}, which occurs when ρ_{12} is high, the variance in the second equation is low and y_{1T+1} is relatively important in the observed second equation (a low a_{12}^2). It is possible that using a sufficiently poor forecast of y_{1T+1} that updating assigns to its best guess of the forecast error ϵ_{1T+2}, that a worse forecast of y_{2T+2} can emerge under updating.

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 information set* $\{y_{11}, y_{12}, ..., y_{1T+1}, ..., y_{1T+s}, y_{21}, ..., y_{2T}\}$, *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 for* $MSE^u < MSE$. *Then the k-step-ahead forecast mean squared error matrix in the updated bivariate VAR is*

$$MSE^u[y_{2T}(k)] = (\sum_{i=0}^{k-1} a_{22}^{2i})(1-\rho_{12}^2)\sigma_2^2 \qquad 2 \leq k \leq s. \tag{2.8}$$

Proof. *See Appendix.*

Clearly, the stationary condition $|a_{22} < 1|$ implies that the MSE of the updated bivariate VAR forecast will converge to the MSE of the ordinary bivariate VAR forecast when the forecast horizon is large enough. Employing the updated bivariate VAR forecast provides a permanent efficiency gain.

2.1.3 Updating Multivariate VAR Forecasts

In the last two sections, we analyzed the updated bivariate VAR forecasts. We now extend this investigation to the multivariate case, where a sequence of informational releases for the variables can occur.

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. The multivariate VAR model (1) is

$$\begin{aligned} y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + ... + a_{1N}y_{Nt-1} + \epsilon_{1t} \quad & t = 1, ..., T+1 \\ y_{2t} &= a_{21}y_{1t-1} + a_{22}y_{2t-1} + ... + a_{2N}y_{Nt-1} + \epsilon_{2t} \quad & t = 1, ..., T \\ & \quad ... & t = 1, ..., T \\ y_{Nt} &= a_{N1}y_{1t-1} + a_{N2}y_{2t-1} + ... + a_{NN}y_{Nt-1} + \epsilon_{Nt} \quad & t = 1, ..., T. \end{aligned} \tag{2.9}$$

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

$$MSE^u[y_{nT}(1)] = (1 - \rho_{1n}^2)\sigma_n^2 \qquad n \geq 2. \tag{2.10}$$

Notice that the MSE for each variable depends on the magnitude of its variance innovation but also the correlation of the innovation with the released variables innovation.

More Available Real-time Variables

We now move on to derive the updated multivariate VAR forecast with m more available real-time variables. First, we consider the $m = 2$ case, that is two variables release data concurrently 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 VAR (1) process

$$\begin{aligned}
y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + ... + a_{1N}y_{Nt-1} + \epsilon_{1t} & t = 1, ..., T+1 \quad (2.11)\\
y_{2t} &= a_{21}y_{1t-1} + a_{22}y_{2t-1} + ... + a_{2N}y_{Nt-1} + \epsilon_{2t} & t = 1, ..., T+1 \quad (2.12)\\
y_{3t} &= a_{31}y_{1t-1} + a_{32}y_{2t-1} + ... + a_{3N}y_{Nt-1} + \epsilon_{3t} & t = 1, ..., T\\
& \quad ... & t = 1, ..., T\\
y_{Nt} &= a_{N1}y_{1t-1} + a_{N2}y_{2t-1} + ... + a_{NN}y_{Nt-1} + \epsilon_{Nt} & t = 1, ..., T.
\end{aligned}$$

We use the known information set $\{y_{11}, ..., y_{1T}, y_{1T+1}; y_{21}, ..., y_{2T}, y_{2T+1}; y_{31}, ..., y_{3T}; ...; y_{N1}, ..., y_{NT}\}$ to forecast $y_{3T}(1)$, ..., $y_{NT}(1)$. Equations (2.11) and (2.12) give

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

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

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

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

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

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

We see that, everything else equal, the weight on the observed innovation is higher the lower is its variance relative to the one forecasted, and the higher the innovation correlation is with the forecasted innovation. 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

$$MSE^u[y_{3T}^{1:2}(1)] = Var[y_{3T+1} - y_{3T+1|T}] = (1 - \frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1-\rho_{12}^2})\sigma_3^2, \quad (2.13)$$

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

This theoretical framework can apply for not only more available real-time variables, but also incorporating consensus data. In general, consensus forecasts are available before the first announcement of many variables. There are situations in which investors must pay for access to expert forecasts and this kind of forecasts are not generally available. If we observe sufficient data on a consensus forecast, then we can use the same updating scheme to improve the forecast of the remaining variables. For example, we can use consensus as one of above two available real-time variables. Then the rest of the updating forecast MSE analysis maintains the same updating scheme. The idea is to treat the consensus variable as any other in the VAR so that the news in consensus will be the value of the released consensus variable minus its predicted value.

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

$$\begin{aligned}
y_{1t} &= a_{11}y_{1t-1} + a_{12}y_{2t-1} + ... + a_{1N}y_{Nt-1} + \epsilon_{1t} & t = 1, ..., T+1 \quad (2.14)\\
&\;\; ... & t = 1, ..., T+1\\
y_{mt} &= a_{m1}y_{1t-1} + a_{m2}y_{2t-1} + ... + a_{mN}y_{Nt-1} + \epsilon_{mt} & t = 1, ..., T+1 \quad (2.15)\\
y_{m+1,t} &= a_{m+1,1}y_{1t-1} + a_{m+1,2}y_{2t-1} + ... + a_{m+1,N}y_{Nt-1} + \epsilon_{m+1,t} & t = 1, ..., T\\
&\;\; ... & t = 1, ..., T\\
y_{Nt} &= a_{N1}y_{1t-1} + a_{N2}y_{2t-1} + ... + a_{NN}y_{Nt-1} + \epsilon_{Nt} & t = 1, ..., T.
\end{aligned}$$

Proposition 4 *Given the information set* $\{y_{11}, ..., y_{1T}, y_{1T+1}; ...; y_{m1}, ..., y_{mT}, y_{mT+1}; y_{m+1,1}, ..., y_{m+1,T}; ...; y_{N1}, ..., y_{NT}\}$, *suppose the random vector* $\epsilon_i = (\epsilon_1 \;\; \epsilon_2 \;\; ... \;\; \epsilon_m)'$ *and the random vector* $\epsilon_j = (\epsilon_{m+1} \;\; \epsilon_{m+2} \;\; ... \;\; \epsilon_N)'$ *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*

$$\begin{aligned} MSE^u[y_{jT}^{1:2:...:m}(1)] &= Var(y_{jT+1} - y_{jT+1|T}) \\ &= Var(\epsilon_{jT+1} - \Omega_{\epsilon_j\epsilon_i}\Omega^{-1}_{\epsilon_i\epsilon_i}\epsilon_{iT+1}) \\ &= Var(\epsilon_{jT+1}) + Var(\Omega_{\epsilon_j\epsilon_i}\Omega^{-1}_{\epsilon_i\epsilon_i}\epsilon_{iT+1}) \\ & \quad -2Cov(\epsilon_{jT+1}, \Omega_{\epsilon_j\epsilon_i}\Omega^{-1}_{\epsilon_i\epsilon_i}\epsilon_{iT+1}). \end{aligned} \tag{2.16}$$

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

Proof. *See Appendix.*

The situation where more than one variable is released will always result in a lower MSE than the case when only one is released. The intuition for the larger fall in MSE is that the additional released variable should be highly correlated with the forecast variable's innovation but have a low correlation with the other released variable's innovation. Essentially, the high innovations should be globally correlated with the forecasted innovations, but not with each other.

Given the information set $\{y_{11}, y_{12}, ..., y_{1T+1}, y_{21}, ..., y_{2T}, ..., y_{N1}, ..., y_{NT}\}$, the forecasting MSE of variable y_{3T+1} by equation (2.10) is

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

However, given the information set $\{y_{11}, ..., y_{1T}, y_{1T+1}; y_{21}, ..., y_{2T}, y_{2T+1}; y_{31}, ..., y_{3T}; ...; y_{N1}, ..., y_{NT}\}$, the forecasting MSE of variable y_{3T+1} is as in (2.13). Comparing (2.10) with (2.13), 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 (2.13) is smaller than the result of equation (2.10). 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, $MSE^u[y_{3T}^{1:2}(1)] < MSE^u[y_{3T}^{1}(1)]$. Clearly, this result extents as we add more released variables.

Proposition 5 *Given the information set* $\{y_{11}, ..., y_{1T}, y_{1T+1}; ...; y_{m1}, ..., y_{mT}, y_{mT+1}; y_{m+1,1}, ..., y_{m+1,T}; ...; y_{N1}, ..., y_{NT}\}$ *and the updated multivariate VAR model (2.14), the more pieces of available information we use, the smaller the forecast mean squared errors. In other words, for a given variable j*

$$MSE[y_{jt}^1(1)] \geq MSE[Y_{jt}^{1:2}(1)] \geq ... \geq MSE[Y_{jt}^{1:...:m-1}(1)] \geq MSE[Y_{jt}^{1:...:m}(1)] \quad (2.17)$$

Proof. *See Appendix.*

2.1.4 Sequential Arrival of Real-time Variables

A natural question arises in considering the sequential release of information: is there an ordering of announcement for firms to obtain maximum disclosure. The answer to this question depends, of course, on what criterion one chooses to assess optimality and what rules exist for the release of information. We set up a simple exercise that we think illustrates the important issues. For example, if all firms released their dates together, then this would provide the most disclosure to all market participants. We think the economic theory as to why earnings disclosures happen over intervals is a worthwhile question, but one outside the scope of this thesis. Indeed, suppose that each firm releases its information for the previous period separately; no two firms can release on the same day (a simplifying assumption that allows for a limiting potential ordering).

The objective function we adopt selects an order of disclosure (enforcing a separate release date) to minimize a standardized mean squared forecast error at each informational release (maximum total disclosure). We assume that the time between releases is sufficiently short that no discounting of future earnings errors is required. Since the VAR structure is itself not related to this criterion, we can calculate for any given set of firms N, any order we wish (including the ordering that satisfies the maximized disclosure), 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 periods of $T + s$, where $T + s$ defines as the number of observations from T to do the forecast objective function. The order is assumed to be fixed over the $T + s$ periods to make the problem meaningful. Otherwise, the order would be chosen on the basis of what firm has the smallest forecast error at that point in time. Also in practice, while order does change, there is typically less variation than would be associated with random draws. Again the persistence in order is an area of research worth pursuing in extensions (see Conclusions).

We minimize the criterion function for all $T+s$, so with N firms there are $N!$ (that is, $N \times (N-1) \times (N-2) \times ... \times 2 \times 1$) different permutations or orderings. Assume that each firm's earnings are normalized in some way (earnings per share or divide them by their standard deviation in order to assure large bonds from discounting) and denote these as $(y_{1t}, y_{2t}, \ldots, y_{nt})$.

Consider the following criterion at release date $T + s$, which we minimize. Given the large number of potential orderings to search over, the notational set-up is somewhat cumbersome. Let the vector $Y_t(i)$ index the i^{th} ordering of informational release (there are $N!$). As we have shown, we will assume no two firms can choose the same day to release information. The i designates a specific ordering of information release. Corresponding to each informational ordering release is an updated forecast depending on the order of release. We illustrate the optimization problems with $N = 3$ which is the simplest example but will demonstrate the key issues for more complicated cases.

To aid in description, let $N = 3$ with the following sequence of informational release sequence: $Y_t(1) = (y_{1t}, y^1_{2t}, y^1_{3t}, y^{1:2}_{3t})$ indicates that Firm 1 releases first, followed by Firm 2 and finally Firm 3, where the superscript 1 indicates that Firm 1 has announced its earnings and the superscript $1 : 2$ indicates that Firm 1 and Firm 2 have released their earnings hereafter; $Y_t(2) = (y_{2t}, y^2_{1t}, y^2_{3t}, y^{2:1}_{3t})$ indicates that Firm 2 releases first, followed by Firm 1 and finally Firm 3; $Y_t(3) = (y_{1t}, y^1_{3t}, y^1_{2t}, y^{1:3}_{2t})$ indicates that Firm 1 releases first, followed by Firm 3 and finally Firm 2; $Y_t(4) = (y_{3t}, y^3_{1t}, y^3_{2t}, y^{3:1}_{2t})$ indicates that Firm 3 releases first, followed by Firm 1 and finally Firm 2; $Y_t(5) = (y_{2t}, y^2_{3t}, y^2_{1t}, y^{2:3}_{1t})$ indicates that Firm 2 releases first, followed by Firm 3 and finally Firm 1; $Y_t(6) = (y_{3t}, y^3_{2t}, y^3_{1t}, y^{3:2}_{1t})$ indicates that Firm 3 releases first, followed by Firm 2 and finally Firm 1. Therefore, there are six possible different orderings (3!).

The forecast error for the first firm is zero and earnings for the remaining two firms are calculated by the updating formula for the VAR using the earnings information from Firm 1. Each of these produces a forecast error $(y_{2t+1} - y^1_{2t+1|t})$ and

$(y_{3t+1} - y^1_{3t+1|t})$. 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 $(y_{3t+1} - y^{1:2}_{3t+1|t})$. Thus, the sum of squared forecast errors for this ordering $Y_t(1)$ is

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

where s denotes the number of forecasting periods. We see that Firm 3 contributes twice to the squared forecast in $S^{1:2}$. The notion here is to derive a single period for information release into N parts. In this case with $N = 3$, once one firm releases data, one of the two remaining firms will delay a second interval finally to release its data. The two period implies a "cost" which is accrued twice. The actual accrual cost depends on the previous ordering. In this case there are five other possible squared forecast sequences: $S^{1:3}, S^{2:1}, S^{2:3}, S^{3:1}, S^{3:2}$, from which 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.

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 (2.10), 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 (2.13), 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

$$\begin{aligned} S^{1:2} &= \sum_{t=T}^{T+s} \frac{1}{s} \left\{(y_{2t+1} - y^1_{2t+1|t})^2 + (y_{3t+1} - y^1_{3t+1|t})^2 + (y_{3t+1} - y^{1:2}_{3t+1|t})^2\right\} \\ &= (1 - \rho^2_{12})\sigma^2_2 + (1 - \rho^2_{13})\sigma^2_3 + (1 - \frac{\rho^2_{13} + \rho^2_{23} - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho^2_{12}})\sigma^2_3. \end{aligned} \tag{2.18}$$

The remaining combinations can be calculated in a similar manner.

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 of the objective function, the informational ordering is Firm 1, followed by Firm 2, and finally Firm 3. In other words, the firm with the largest correlation to other firms releases its earnings earlier in order to obtain the minimum mean squared forecast error.*

Proof. *See Appendix.*

Holding everything else equal, firms whose innovations are highly correlated release first. The intuition is straightforward. Since firms with higher correlation coefficients will provide the most useful information for forecasting the remaining data.

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. The firm with the largest variance releases its information first in order to obtain the minimum mean squared forecast error.*

Proof. *See Appendix.*

2.2 Contemporaneous Ordinary Least Squares Regression

A natural way to estimate the current earnings of one firm is to use current earnings of firms' data that have been released. That is, use historical contemporaneous data in a regression and update the forecast with the estimated coefficients and any released data. In this section we study the prediction properties with a simple contemporaneous regression model.

In a simple contemporaneous regression model, suppose that y_1 is observable from 1 to $T+1$ and y_2 is observable from 1 to T. If our goal is to predict y_2 at $T+1$, then we can simply regress y_2 on y_1 and the forecast of y_2 at $T+1$ is the coefficient estimate multiplied by the known value (released) of y_1 at $T+1$. Since the variables are determined contemporaneously, but just differ as it when the information is released, one would expect this prediction to be based on biased estimating.

The simple contemporaneous model is

$$y_{2t} = \beta y_{1t} + \epsilon_t. \tag{2.19}$$

Suppose the true DGP is a bivariate VAR (1) process

$$y_{1t} = a_{11}y_{1t-1} + a_{12}y_{2t-1} + \epsilon_{1t} \tag{2.20}$$

$$y_{2t} = a_{21}y_{1t-1} + a_{22}y_{2t-1} + \epsilon_{2t}. \tag{2.21}$$

We see from (2.20) that $y_{2t-1} = (y_{1t} - a_{11}y_{1t-1} - \epsilon_{1t})/a_{12}$. If we substitute this into

(2.21), we find that

$$y_{2t} = \frac{a_{22}}{a_{12}} y_{1t} + (a_{21} - \frac{a_{11}a_{22}}{a_{12}}) y_{1t-1} + \epsilon_{2t} - \frac{a_{22}}{a_{12}} \epsilon_{1t}$$
$$= \beta y_{1t} + \delta y_{1t-1} + u_t, \quad (2.22)$$

where

$$\delta \equiv a_{21} - \frac{a_{11}a_{22}}{a_{12}}, \quad \beta \equiv \frac{a_{22}}{a_{12}}, \quad u_t \equiv \epsilon_{2t} - \beta\epsilon_{1t}.$$

Thus $Var(u_t)$ is equal to $\sigma_2^2 + \beta^2\sigma_1^2 - 2\beta\rho_{12}\sigma_1\sigma_2$.

Since y_{1t} is correlated to the error term ϵ_{1t} by assumption, it is also correlated to the error term u_t. Here ϵ_{1t} is the measurement error which is assumed to be identified, independent, and distributed with variance σ_1^2 and to be independent of y_{1t-1} and y_{2t}. It is also assumed that there is contemporaneous correlation between ϵ_{1t} and ϵ_{2t}. Therefore, $E(\epsilon_{1t}\epsilon_{2t}) = \rho_{12}$ for some correlation coefficient ρ_{12} such that $-1 < \rho_{12} < 1$.

If we estimate a simple model like (2.19) with $\delta = 0$, one effect of the measurement error in the independent variable is to increase the variance of the error terms if ϵ_1 and ϵ_2 are negatively correlated. Another severe consequence is that the OLS estimator is biased and inconsistent. Because u_t depends on ϵ_{1t}, u_t must be correlated with y_{1t} whenever $\beta \neq 0$. In fact, since the random part of y_{1t} is ϵ_{1t} and ϵ_{1t} is correlated to ϵ_{2t}, we have that

$$E(u_t|y_{1t}) = E(u_t|\epsilon_{1t}) = \epsilon_{2t} - \beta\epsilon_{1t}.$$

Using the fact that $E(u_t) = 0$ unconditionally, we can see that

$$\begin{aligned} Cov(y_{1t}, u_t) &= E(y_{1t}u_t) \\ &= E(y_{1t}E(u_t|y_{1t})) \\ &= E((y_{1t})(\epsilon_{2t} - \beta\epsilon_{1t})) \\ &= \rho_{12}\sigma_1\sigma_2 - \beta\sigma_1^2. \end{aligned}$$

This covariance is negative if $\beta > \rho_{12}\sigma_2/\sigma_1$ and positive if $\beta < \rho_{12}\sigma_2/\sigma_1$. Since this covariance does not depend on the sample size T, it does not go away as T becomes large. Therefore the OLS assumption that $E(u_t|X_t) = 0$ is false whenever any element of X_t, that is y_{1t} and y_{1t-1} in our special case, is measured with error. In consequence, the OLS estimator is biased and inconsistent.

Proposition 8 *Given the true DGP of VAR (1) in equations (2.20) and (2.21), the OLS estimator in the contemporaneous regression model (2.19) is biased and inconsistent.*

Proof. *See Appendix.*

This proposition shows when the true DGP is a bivariate VAR (1), the contemporaneous regression model is misspecified and OLS estimators lead to a serious measurement error problem. Since the two variables are jointly determined, the contemporaneous regression model can result in either an over- or under-estimate.

2.3 Conclusions

In VAR time series models, with variables released along a sequence, the covariance matrix of the innovations plays an important role in updating forecasts. We propose a practical method to update forecasts in multivariate VAR models. The focus and the benefit of employing the updating approach is that the innovations of the currently known observations are useful in predicting innovations of the to-be-released observations and hence can lead to more accurate forecasts. The theoretical framework shows that the currently known observations of one variable will be useful for forecasting the currently unknown observations of other variables so long as the innovations are correlated. Higher correlation among observation innovations of multi-variables implies that the mean squared forecast error of the currently unknown observations of the other variables will be more smaller than the usual.

We first derive the general mean squared error of a multistep-ahead forecast in an updated bivariate VAR, which is almost always smaller than the corresponding standard VAR forecast. The updated VAR model is then used to perform a long-horizon forecast to assess the forecast mean squared errors. We also 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 a flu outbreak or a vaccine failure, motivates this kind of exercise.

In specifying VARs, we choose a series of related variables comprising the system. The usual aim is to have the lags of own variables as well as the cross ones. Innovation correlation, when all the variables are the same in each equation, are not important for efficient estimation. However, valid inference requires that the covariance of innovation is taken into account. In this forecasting context, new information is in the form of innovations that can update guesses on all other innovations for variables not yet released. The key of course is that innovations need to be correlated.

Chapter 3

An Application to U.S. Nonfarm Payroll Employment

3.1 Introduction

Federal government agencies regularly announce the latest calculated values for economic variables. A monthly announcement reports the series' value in the last month. As well, firms announce their quarterly financial report (for instance, earnings per share and net income) for public use. The schedule for these announcements is known well in advance, generally by the previous year-end. The timing over the interval of these announcement either varies or is fixed. The order in which variables are announced may also vary or be fixed each month or quarter.

Although there are numerous macroeconomic announcement series or earnings announcements for firms that we could consider to apply the VAR updating method, we choose one of the real macroeconomic variables, namely, nonfarm payroll employment from the Bureau of Labor Statistics (BLS) in the United States, as an excellent candidate first for updating. It has been used in the formulation of Federal Reserve policy, and is widely watched by market participants. In addition, nonfarm payroll employment announcements have fixed announcement timing and order.

One employment announcement used in the bivariate VAR model is total nonfarm payroll employment. The BLS, U.S. Department of Labor, releases the employment data each month. The announcements are usually made at 8:30 a.m. on a Friday. The employment information is composed of household survey and establishment survey data. The household survey has a wider scope than the establishment survey since it includes the self-employed, unpaid family workers, agricultural workers, and private household workers, who are excluded from the establishment survey. However, the establishment survey employment series has a smaller margin of error on the measurement of month-to-month change than the household survey, which has a much larger sample size. The establishment survey includes payroll employment information, such as total nonfarm payroll employment, weekly and hourly earnings, and weekly hours worked for several industries. As a set of labor market indicators, nonfarm payroll employment counts the number of paid employees working part-time or full-time in the nation's business and government establishments.

Another employment announcement which precedes the BLS release is private nonfarm payroll employment. Automatic Data Processing (ADP) contracted with Macroeconomic Advisors to compute a monthly report (ADP National Employment Report). Estimates of employment published in the ADP National Employment Report were made available beginning in January of 2001. Most of the announcements are made at 8:15 a.m. on a Wednesday (the previous Wednesday to the BLS Friday release), although a few announcements are made on other days, but always prior to BLS release. All announcement dates, whether Wednesday or not, are included in our study. The ADP report is a measure of nonfarm private employment, and it calculates the level of employment by select industry and by size of payroll (small, medium, and large). It ultimately helps market participants to use this informally to adjust their prediction of monthly nonfarm payrolls from the BLS. The ADP report only covers private payrolls, excluding government and has a smaller coverage than the BLS. Nevertheless, nonfarm private employment released on the behalf of the ADP (hereafter called ADP's data) is highly correlated with nonfarm payroll employment announced two days later by the BLS (hereafter called BLS's data). The correlation over the sample is 0.82 for the change in each variable.

Table 3.1: Descriptive Statistics of Changes in Stock Price, ADP, and BLS

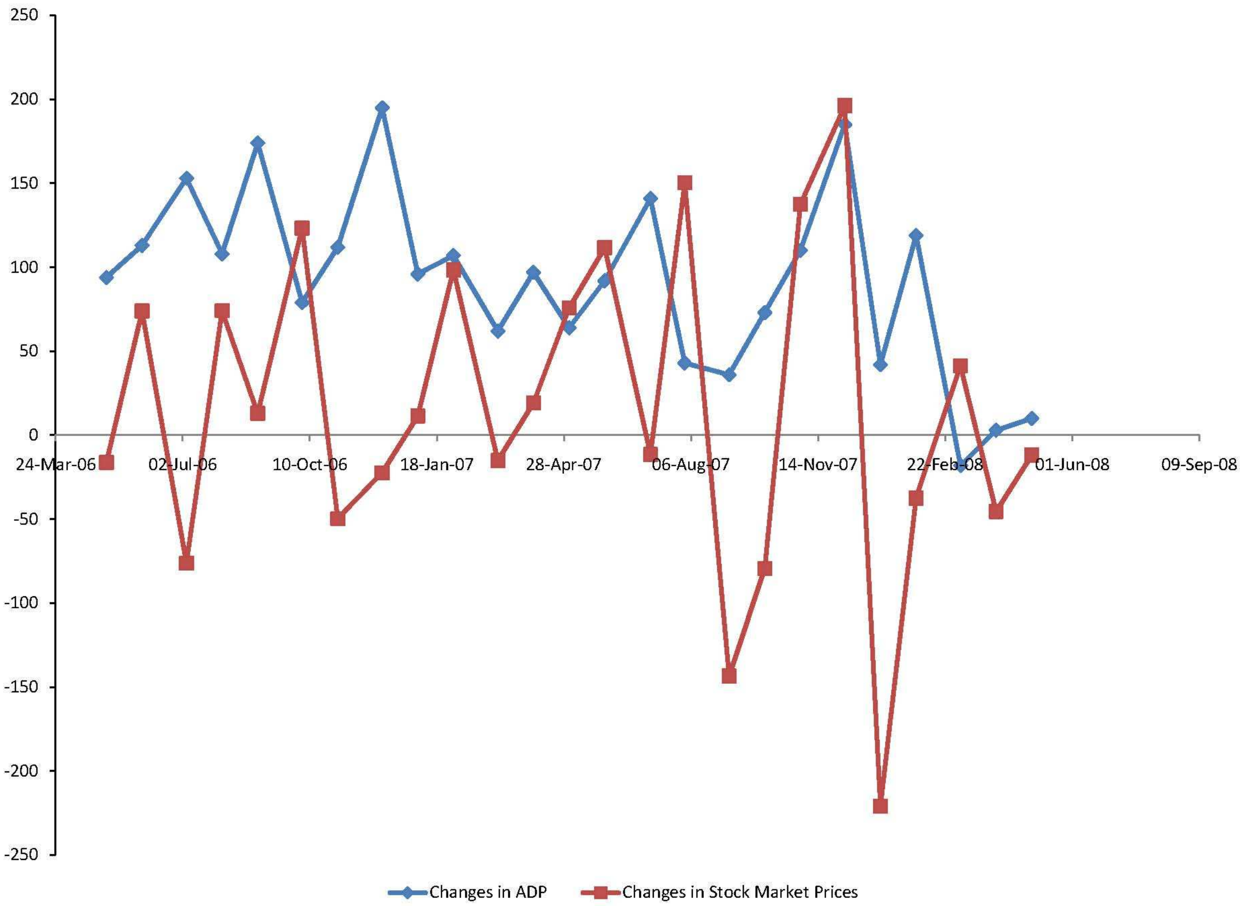
	Δ ADP	Δ Stock Price†	Δ BLS	Δ Stock Price*
Mean	91.6	15.87	86.84	-31.43
S.D	54.59	95.48	91.34	108.63
Corr	0.22		0.18	

†Changes in stock market price is the differences of daily closing price in Dow Jones Index between the monthly ADP announcement date and one day prior to the monthly ADP announcement date.
* Changes in stock market price is the differences of daily closing price in Dow Jones Index between the monthly BLS announcement date and one day prior to the monthly BLS announcement date.

Total nonfarm payroll employment by the BLS is the first major economic indicator released each month. To gain some sense of its importance, we illustrate the correlations between stock market prices and both changes in private nonfarm payroll employment and changes in total nonfarm payroll employment in Table 3.1. We employ the differences of daily closing price in Dow Jones Index between the monthly ADP (or BLS) announcement date and one day prior to the monthly ADP (or BLS) announcement date as changes in stock market prices. Figure 3.1 illustrates the correlation between ADP announcements and stock market prices at the date of the ADP announcement. The mean of changes in private employment and changes in stock price are 91.6 and 15.87 respectively. The standard deviation of changes in private employment and changes in stock price are 54.59 and 95.48 respectively. The correlation between these changes is 0.22. Figure 3.2 illustrates correlation between BLS announcements and stock market prices at the date of the BLS announcement. The mean of changes in total employment and changes in stock price are 86.84 and -31.43 respectively. The standard deviation of changes in total employment and changes in stock price are 91.34 and 108.63 respectively. The correlation between these changes is 0.18. Figure 3.3 gives the full picture of quantitative co-movement between the first release of the BLS employment information and the Dow Jones industrial index closing price. It illustrates the growth rate of the monthly BLS first release, the final release, and the daily Dow Jones Index closing price. For instance, on September 7, 2007, the BLS reported that in August, the number of nonfarm payroll employment decreased by 4,000. This led the stock market index to a decline by 250 points in the closing price. On Nov 2, 2007, the BLS reported a 166 thousand increase in nonfarm employment. This led the stock index closing price to rise by 28 points on that day.

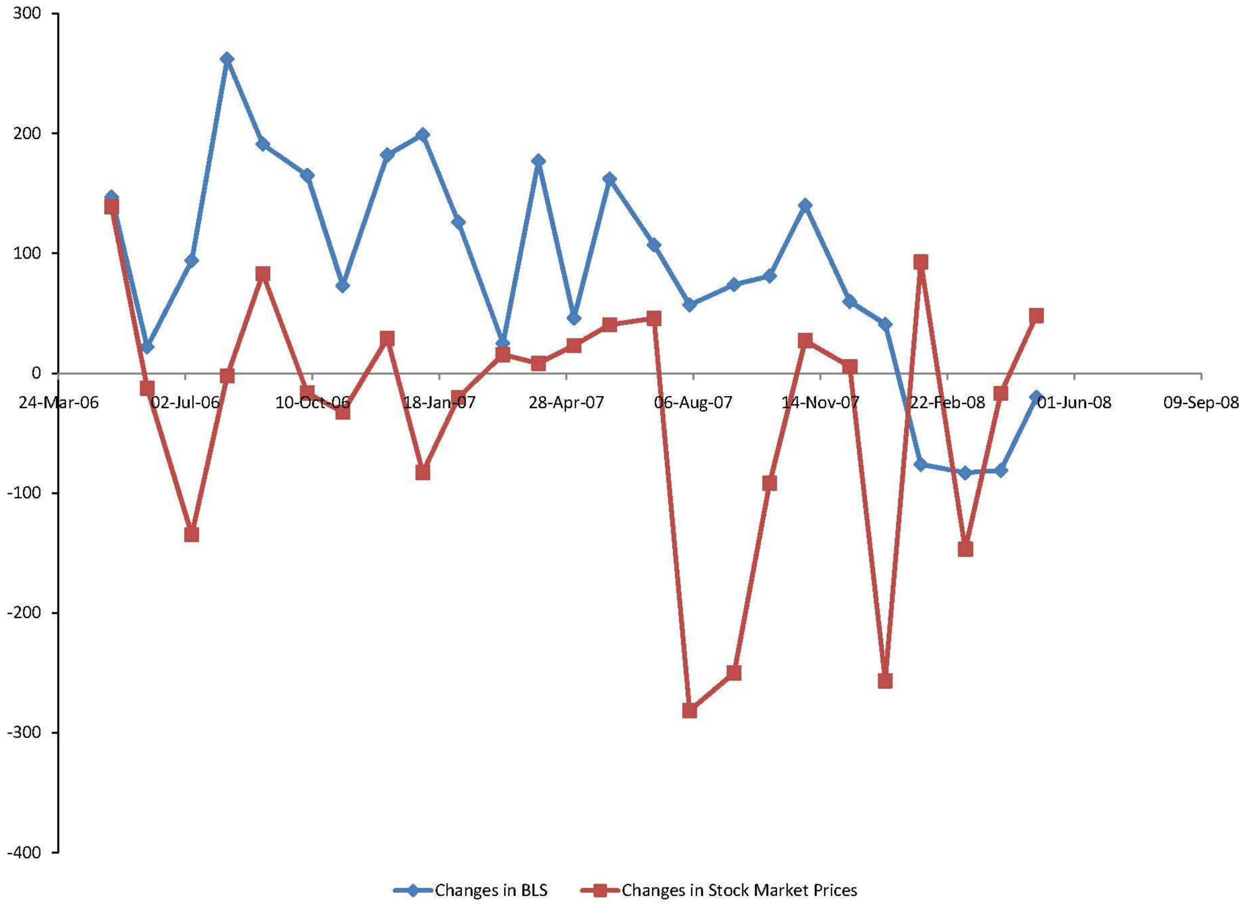
Both the ADP and the BLS are subject to substantial revisions as new information becomes available. The ADP revises its initial release one month later. In addition to monthly revision, the ADP recomputes estimates of private nonfarm employment on an annual basis. The BLS revises its initial monthly estimates twice in the immediately succeeding two months. On an annual basis, the BLS recalculates estimates to complete employment counts available from unemployment insurance tax records. A

Figure 3.1: Correlation of Private Employment Announcements on Stock Market



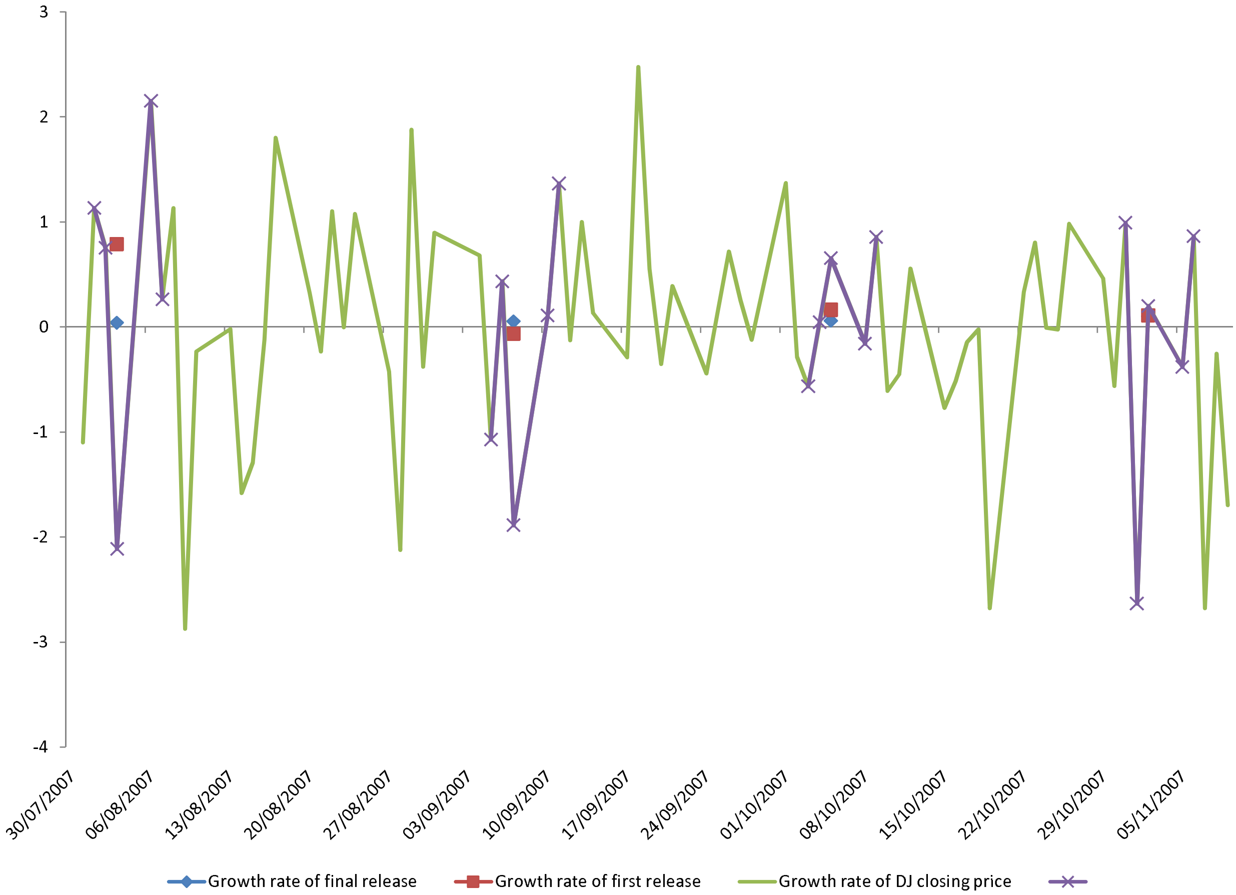
The mean of changes in private employment and changes in stock price are 91.6 and 15.87 respectively. The standard deviation of changes in private employment and changes in stock price are 54.59 and 95.48 respectively. The correlation between changes in private employment and changes in stock price is 0.22.

Figure 3.2: Correlation of Total Employment Announcements on Stock Market



The mean of changes in total employment and changes in stock price are 86.84 and -31.43 respectively. The standard deviation of changes in total employment and changes in stock price are 91.34 and 108.63 respectively. The correlation between changes in total employment and changes in stock price is 0.18.

Figure 3.3: Correlation of the First and Final Employment Announcements on Stock Market



The vertical axis represents the growth rate of monthly BLS first release, final release, and daily Dow Jones Index closing price.

common question in forecasting is: which data to use? Should we use the first released data to make a real-time forecasting? Or should we use the final revised data to take into account unemployment insurance tax records? Since market participants tend to react to the first release with subsequent revision being largely ignored, investors can take not only more strategic control of their portfolio, but also advantage of unique investment opportunities that arise in the days surrounding the BLS report. Also, our updating forecast method is short-lived (until the release of all data). Therefore, in this study we use the data at the time of forecasting to predict first release values of the BLS data. The real-time data sets we create are updated in the data set, in principle, whenever the data source is available in real-time. In this series, data sets change (mostly because of recent revision) for each forecast, so careful attention is required to use data settlement that are publicly available when forecasts are made.

3.2 Data

The ADP's data (adp) and BLS's data (bls), are used to fit the bivariate VAR model in (2.2). The 101 observations are seasonally adjusted monthly from January 2001 to May 2009. As mentioned before, for each of the forecasting exercises, the data used are those available at the time the forecast was made. Table 3.2 reports the descriptive statistics for the first difference (levels are nonstationary and most market watchers follow changes). Dickey-Fuller nonstationary tests have been conducted, and the presence of a unit root is rejected. Both series are stationary with first differences. Since 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 3.4. After the first difference

Table 3.2: Descriptive Statistics

Source	Automatic Data Processing, Inc. Macroeconomic Advisers, LLC.B Bureau of Labor Statistics	
Frequency	Seasonally adjusted, Monthly	
Sample period	January 2001 - May 2009	
Sample Size	101	
	1^{st} difference of *adp*	1^{st} difference of *bls*
Mean	-15.13	-3.31
Standard deviation	222.27	240.00
β_1	-0.06	-0.13
Augmented D-F	-1.6	-2.53
H_0 : Nonstationarity	Reject at 90% (critical value = -1.6)	

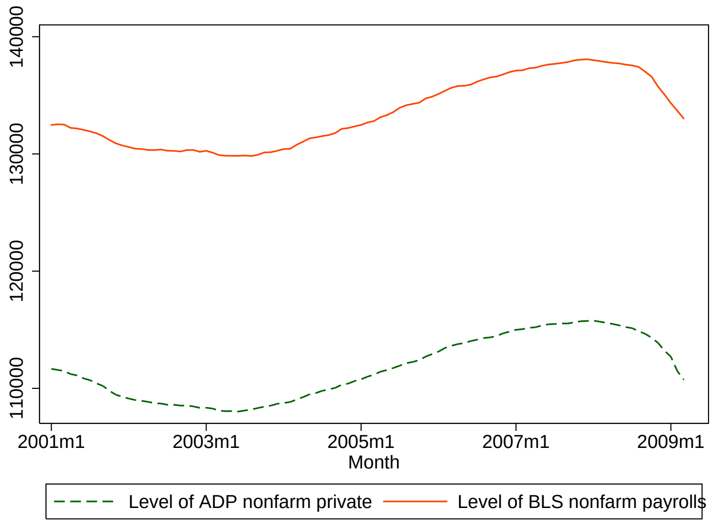
The coefficient β designates the autocorrelation of the series at lag i. The augmented Dickey-Fuller test is based on the following regression: $y_t - y_{t-1} = \beta_0 + \beta_1 y_{t-1} + \beta_2 \Delta y_{t-1} + \beta_3 \Delta y_{t-2} + \mu_t$. Terms added until additional lags provide no new information significant at the 10% level.

of each time series, the plot of the monthly changes of employees on nonfarm payrolls is provided in Figure 3.5.

The timing of the release of the two time series is fixed and in order of the ADP's release first, followed by the BLS. Table 3.3 shows that the ADP national employment report is released, for public use only, two days prior to publication of employment information by the BLS. Due to official holidays, the June and August 2008 ADP national emplyment announcements are ones which were released for public use only one day prior to publication of employment by the BLS. The time table below constitutes the sample period for our forecasting exercise. Data revisions are incorporated as they occur.

The BLS revises its initial monthly estimates twice, in the immediately succeeding two months, to incorporate additional information that was not available at the time

Figure 3.4: Level of Nonfarm Payrolls



Source: Automatic Data Processing, Inc., Macroeconomic Advisers, LLC., Bureau of Labour Statistics. Nonfarm payroll employment counts the number of paid employees working part-time or full-time in the nation's business and government establishments. The ADP national employment report only covers private payrolls, excluding government.

of the initial publication of the estimates. On an annual basis, the BLS recalculates estimates to complete employment counts available from unemployment insurance tax records, usually in March. ADP revises its initial release every one month later. In addition to monthly revision, the entire history of the estimate of private nonfarm employment covering the period from January 2001 through January 2007 was revised on Thursday, February 22, 2007. The first regular release of the ADP national employment report to incorporate all these revisions was published on Wednesday, March 7, 2007.

We incorporate the features of revisions existing in the *bls* and *adp* time series into time series model estimation and forecasting. Rather than using finally released data

Figure 3.5: Changes in Nonfarm Payrolls

to estimate and forecast, we employ data of as many different releases as there are dates in the sample. More specifically, at every date within a sample, both right-side and left-side variables in the bivariate VAR model should be the most up-to-date estimates of the variables at that time.

3.3 Forecasting

A simple bivariate VAR is estimated over the base period. Each equation can be estimated by ordinary least squares (OLS) regression. Since each equation contains the same explanatory variables, OLS on each equation is efficient. The number of lagged values to include in each equation is determined by Schwarz's Bayesian information criterion (SBIC). SBIC suggests the optimal lag selection of one. We employ one lag in all time series models such as the updated VAR and the ordinary VAR.

Table 3.3: U.S. Nonfarm Payrolls Employment Announcements

	2007 Release Date		2008 Release Date	
Reference Month	ADP	BLS	ADP	BLS
January	Jan.31, Wed.	Feb.2, Fri.	Jan.30, Wed.	Feb.1, Fri.
February	Mar.7, Wed.	Mar.9, Fri.	Mar.5, Wed.	Mar.7, Fri.
March	Apr.4, Wed.	Apr.6, Fri.	Apr.2, Wed.	Apr.4, Fri.
April	May.2, Wed.	May.4, Fri.	Apr.30, Wed.	May.2, Fri.
May	May.30, Wed.	Jun.1, Fri.	Jun.4, Wed.	Jun.6, Fri.
June	Jul.5, Thu.	Jul.6, Fri.	Jul.2, Wed.	Jul.3, Thu.
July	Aug.1, Wed.	Aug.3, Fri.	Jul.30, Wed.	Aug.1, Fri.
August	Sep.5, Wed.	Sep.7, Fri.	Sep.4, Thu.	Sep.5, Fri.
September	Oct.3, Wed.	Oct.5, Fri.	Oct.1, Wed.	Oct.3, Fri.
October	Oct.31, Wed.	Nov.2, Fri.	Nov.5, Wed.	Nov.7, Fri.
November	Dec.5, Wed.	Dec.7, Fri.	Dec.3, Wed.	Dec.5, Fri.
December	Jan.2, 09, Wed.	Jan.4, 09, Fri.	Jan.7, 09, Wed.	Jan.9, 09 Fri.

The bivariate VAR (1) model estimation with standard deviation under the coefficients over the sample period of January 2001 through May 2009 is as follows:

$$\begin{aligned} adp_t &= \underset{(8.80)}{-8.15} + \underset{(0.12)}{0.68} \ \ adp_{t-1} + \underset{(0.11)}{0.25} \ \ bls_{t-1} + \epsilon_{1t} \\ bls_t &= \underset{(11.74)}{2.10} + \underset{(0.16)}{0.53} \ \ adp_{t-1} + \underset{(0.15)}{0.41} \ \ bls_{t-1} + \epsilon_{2t}. \end{aligned} \tag{3.1}$$

Interestingly, lagged *adp* is more important in predicting *bls* than its own lag. This suggests that knowledge of adp_t for forecasts of horizon longer than two periods will be helpful. The correlation of residuals in the bivariate VAR is 0.72, which is quite high and again implies that is forecasting along linear updating. In the ordinary VAR forecast, the coefficient estimates and the variance covariance matrix estimates of residuals are used to calculate the forecast mean squared error. In the updated VAR forecast, in addition to coefficient estimates and variance covariance matrix estimates of residuals, we also consider the correlation of residuals of the VAR to compute the more efficient forecast mean squared error. Table 3.4 shows variance covariance matrix estimates of residuals.

Table 3.4: Variance Covariance Matrix Estimates of Residuals

	ADP Residuals	BLS Residuals
ADP Residuals	7671.67	
BLS Residuals	7367.99	13659.10

This is a real-time forecast with revisions in the data set. For each time period forecast, we collect all the information available at that time to make one- and two-step-ahead forecasts. To provide the reader an idea of the number and magnitude of the revisions, consider Table 3.5. It shows revisions in the data set and the data set used for estimation in each time period.

Knowing *adp*'s first release is always two days prior to *bls*'s first release, we take advantage of the two-days lead in *adp* and forecast *bls* in the same time period. The detailed estimation method follows the theoretical strategy in Chapter 2. In practical examples, of course, we use estimates of coefficients and residual correlations. The estimation period for reference is January 2001 through June 2007. With the August 1 release of ADP for July 2007, we have all the information for the updated VAR. At the first stage, we estimate the bivariate VAR over the period from January 2001 through June 2007. Following maximum-likelihood estimation by VAR, the one-step-ahead residual prediction of both series is straightforward. At the second stage, we

Table 3.5: Revisions in Data

On	Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS
2007m1	107	347	107	347	107	347	107	347	107	347	107	347	107	347	107	347	107	347
2007m2	62	90	62	90	62	90	62	90	62	90	62	90	62	90	62	90	62	90
2007m3	148	177	148	175	148	175	148	175	148	175	148	175	148	175	148	175	148	175
2007m4	61	88	61	80	61	122	61	122	61	122	61	122	61	122	61	122	61	122
2007m5	97		98	157	98	190	98	188	98	188	98	188	98	188	98	188	98	188
2007m6			150		143	132	143	126	143	69	143	69	143	69	143	69	143	69
2007m7					48		41	92	41	68	41	93	41	93	41	93	41	93
2007m8							38		27	-4	27	89	27	93	27	93	27	93
2007m9									58		61	110	61	96	61	44	61	44
2007m10											106		119	166	119	170	119	159
2007m11													189		173	94	173	115
2007m12															40		37	18
2008m1																	130	
2008m2																		
2008m3																		
2008m4																		
2008m5																		
2008m6																		
2008m7																		
2008m8																		
2008m9																		

On	Jan		Feb		Mar		April		May		June		July		Aug	
	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS	ADP	BLS
2007m1	107	347	107	347	107	126	107	126	107	126	107	126	107	126	107	126
2007m2	62	90	62	90	62	25	62	25	62	25	62	25	62	25	62	25
2007m3	97	175	97	175	97	177	97	177	97	177	97	177	97	177	97	177
2007m4	64	122	64	122	64	46	64	46	64	46	64	46	64	46	64	46
2007m5	92	188	92	188	92	162	92	162	92	162	92	162	92	162	92	162
2007m6	141	69	141	69	141	107	141	107	141	107	141	107	141	107	141	107
2007m7	43	93	43	93	43	57	43	57	43	57	43	57	43	57	43	57
2007m8	36	93	36	93	36	74	36	74	36	74	36	74	36	74	36	74
2007m9	73	44	73	44	73	81	73	81	73	81	73	81	73	81	73	81
2007m10	110	159	110	159	110	140	110	140	110	140	110	140	110	140	110	140
2007m11	185	-325	185	-325	185	60	185	60	185	60	185	60	185	60	185	60
2007m12	42	82	42	41	42	41	42	41	42	41	42	41	42	41	42	41
2008m1	119	-17	119	-22	119	-76	119	-76	119	-76	119	-76	119	-76	119	-76
2008m2	-23		-18	-63	-18	-76	-18	-83	-18	-83	-18	-83	-18	-83	-18	-83
2008m3			8		3	-80	3	-71	3	-88	3	-88	3	-88	3	-88
2008m4					10		13	-20	13	-28	13	-67	13	-67	13	-67
2008m5							40		25	-49	25	-62	25	-47	25	-47
2008m6									-79		-77	-62	-77	-51	-77	-100
2008m7											9		1	-51	1	-60
2008m8													-33		-37	-84
2008m9															-8	

regress the residuals of the *bls* on the residuals of the *adp*. Since we observe the actual error term of the *adp* in July 2007, the best fitted residual of the *bls* in July 2007 is the estimated coefficient multiplied by the actual error term of the *adp* in July 2007.

The updated VAR forecasts requies the estimated correlation over the available sample and the residual of the released variables (*adp*) from what was predicted. This exactly follows the theoretical development except we do not observe the actual innovation since we use the estimated VAR. That is, $\tilde{\epsilon}_{bls2007} = \hat{\rho}\hat{\epsilon}_{adp2007}$. One-step-ahead forecasts for *bls* are then done for July 2007. If we want to do multistep forecasts, we can follow the method in Chapter 2. In the exercise below we consider one-step-ahead (which we have called monthly) and two-step-ahead forecasts.

3.4 Forecast Accuracy Comparison

The BLS releases revision of past employment announcements for the previous three months, after which the announcement is considered final. Because the ultimate test of a forecasting model is its out-of-sample performance, in this section we focus on out-of-sample forecasts. Multistep ahead forecasts are computed by forward iteration of the VAR. We first estimate over the sample period of 2001m1 through 2007m6, and forecast *bls* for 2007m7 (one-step-ahead) and 2007m8 (two-step-ahead); we then estimate over the sample period of 2001m1 through 2007m7, and forecast *bls* for 2007m8 (one-step-ahead) and 2007m9 (two-step-ahead); and so on. We repeat the process until we are through the data set. We then calculate the mean squared errors for the updated VAR forecasting.

As a comparison, out-of-sample forecasts were also computed for a standard vector autoregression model with one lag, that is, a regression of the variable on lags of its own past values and of other variables' past values. The key difference between the updated VAR forecasts and the ordinary VAR forecasts is the first stage estimation. Taking advantage of one more piece of information at the first forecasting period, the updated VAR forecast provides the best estimate we are going to obtain.

A total of 23 observations are constructed by one-step-ahead time series forecasting based on the time period of July 2007 through May 2009. For instance, on August 1, 2007, after ADP reported the first released nonfarm private employment, we use adp from January 2001 through July 2007 and bls from January 2001 through June 2007 to forecast bls in July 2007. On August 1, 2007, it was forecasted that the data for July 2007 would be released on August 3, 2007. We calculate the 23 observations of one-step-ahead forecasts, which are shown in Figure 3.6. The solid line is the BLS first released data from July 2007 through May 2009. The first released data plays a key role in markets, so that we employ it as the actual value. We find that the updated VAR forecast outperforms the ordinary VAR forecast.

The two-step-ahead time series forecast is based on the time period of August 2007 through May 2009. In total, 22 observations are constructed. The difference from one-step-ahead forecast is that based on the one-step-ahead forecast of bls in July 2007, an iterated one-step-ahead forecast of bls in August 2007 is conducted. Since the two-step-ahead forecast is based on the one-step-ahead forecast, the updated multistep VAR forecast outperforms the ordinary VAR forecast. Figure 3.7 reports the updated VAR forecasts and the ordinary VAR forecasts.

Table 3.6 shows the root mean squared forecast error for each of the forecasting methods. The mean squared forecast error is computed as the average squared value of the forecast error over the out-of-sample time period of July 2007 through May

Figure 3.6: One-step-ahead Forecast: 2007m7 — 2009m5

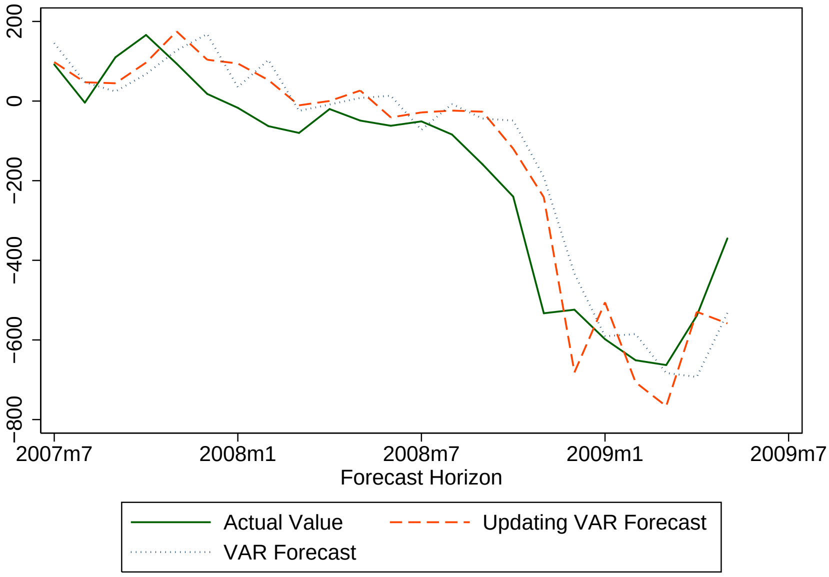
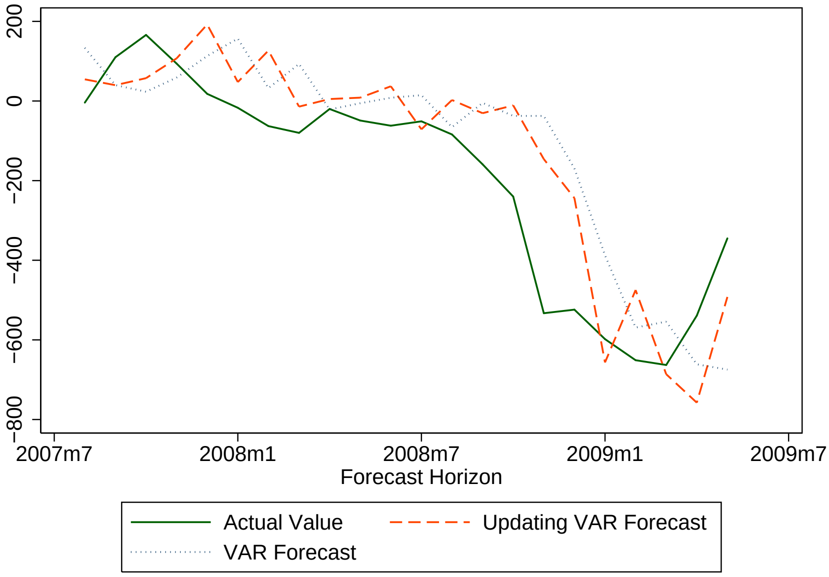


Figure 3.7: Two-step-ahead Forecast: 2007m8 — 2009m5



2009, and the resulting square root is the root mean squared forecast error reported in the table. The entries of column 2 and 3 are the root mean squared errors of the updated VAR forecast and the ordinary VAR forecast, respectively, for nonfarm payroll employment *bls*. The results indicate that the updated VAR forecast has lower root mean squared error than the ordinary VAR forecast over one- and two-step-ahead forecast. Compared with the ordinary VAR, the relative efficiency gain by using the updated VAR forecast is 8% in the one-step-ahead forecast and 16% in the two-step-ahead forecast.

Table 3.6: Root Mean Squared Errors of Out-of-Sample Forecasts

Forecast Horizon	Updating VAR	Ordinary VAR	Efficiency Gain
1 month	109.65	120.42	8%
2 months	153.53	185.60	16%

3.5 How reliable are these confidence intervals

We have discussed a great deal the forecast mean squared error by the updated VAR and the ordinary VAR methods. However, tighter confidence intervals of forecasts are also an important way to see how reliable these forecasts are. In this section, we study the asymptotic and bootstrap confidence intervals for the examples that we did for *bls* of one- and two-step-ahead forecasts.

Using U.S. monthly employment data over the sample periods of 2001m1 – 2007m6, we estimate the bivariate model (3.1) and yield the restricted estimates $\hat{A}$, $\hat{\mu}$, and $\hat{\Omega}$, where

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

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

and $\hat{\mu}$ is the mean zero vector, that is the mean residual of *adp* and the mean residual of *bls*. We then compute one- and two-step-ahead forecasts $f\hat{cast}$ for 2007m7 and forecast standard errors $\hat{s}_{fcast}$ by the updated VAR and the ordinary VAR respectively. The bootstrap DGP is constructed by

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

which is just the model (2.2) 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 t^{th} element of the bootstrap sample, must depend on both y_{1t-1}^* and y_{2t-1}^* simultaneously. Notice that the bootstrap sample is conditional on the observed values of $y_{1,0}$ and $y_{2,0}$, where we set them equal to zero initially. 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

ρ^*_{12}. The bootstrap t-statistic is then calculated by

$$t^*_j = \frac{fcast^* - f\hat{cast}}{\sqrt{(1-\rho^{*2}_{12})}\ s^*_{fcast}}.$$

Consider the $(1-\alpha)$ confidence interval. Let us denote $r_{\alpha/2}$ as the smallest integer not less 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 $f\hat{cast} - \sqrt{(1-\hat{\rho}^2_{12})}\hat{s}_{fcast}\ c^*_{\alpha/2}$. Further, the lower limit of the confidence interval is $f\hat{cast} - \sqrt{(1-\hat{\rho}^2_{12})}\hat{s}_{fcast}\ c^*_{1-(\alpha/2)}$, 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

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

Next, we estimate the bivariate model (3.1) over the sample periods of 2001m1 – 2007m7 and access $\hat{A}$ and $\hat{\Omega}$ again. We then compute one- and two-step-ahead forecasts $f\hat{cast}$ for the time period of 2007m8 and forecast standard errors $\hat{s}_{fcast}$ by the updated VAR and the ordinary VAR respectively. We repeat each bootstrap simulation $R = 23$ times to calculate a bootstrap confidence interval for each forecasting period until we are through the whole dataset.

Figures 3.8 and 3.9 illustrate 99% bootstrap confidence intervals of the updated VAR one-step-ahead forecasts and the ordinary VAR one-step-ahead forecasts respectively. Given 23 observations in the one-step-ahead forecasting time periods of 2007m7 to 2009m5, all 23 actual values fall in the 99% confidence intervals of the updated VAR and the ordinary VAR. Figures 3.10 and 3.11 demonstrate 99% bootstrap confidence intervals of the updated VAR two-step-ahead forecasts and the ordinary VAR two-step-ahead forecasts respectively. Given 22 observations in the forecasting time periods of 2007m8 to 2009m5, there are 22 actual values falling into the

Figure 3.8: Bootstrap Confidence Intervals for One-step-ahead Updated Forecast: 2007m7 — 2009m5

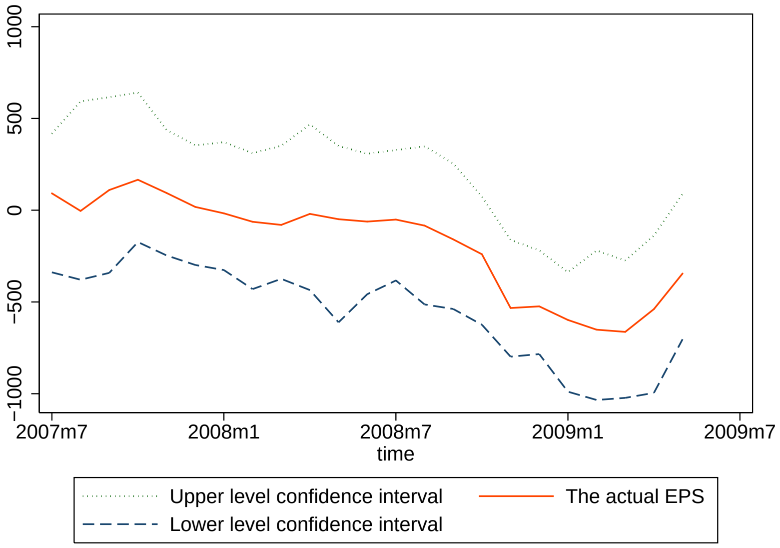


Figure 3.9: Bootstrap Confidence Intervals for One-step-ahead Ordinary Forecast: 2007m7 — 2009m5

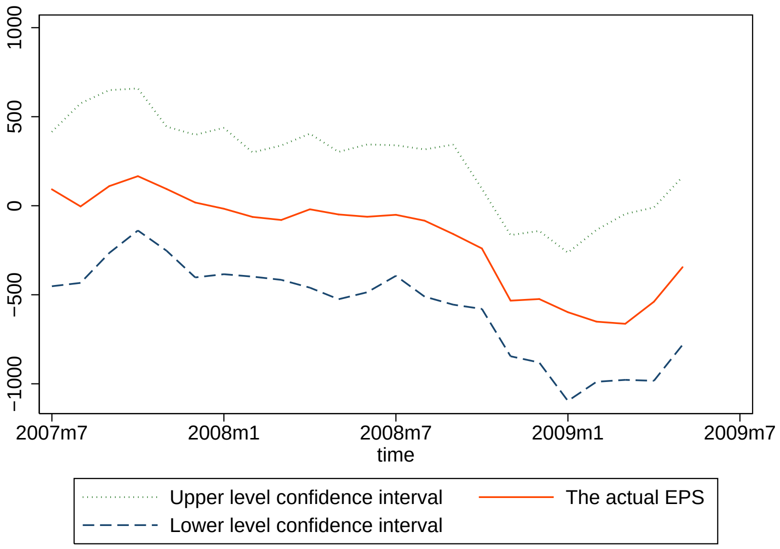


Figure 3.10: Bootstrap Confidence Intervals for Two-step-ahead Updated Forecast: 2007m8 — 2009m5

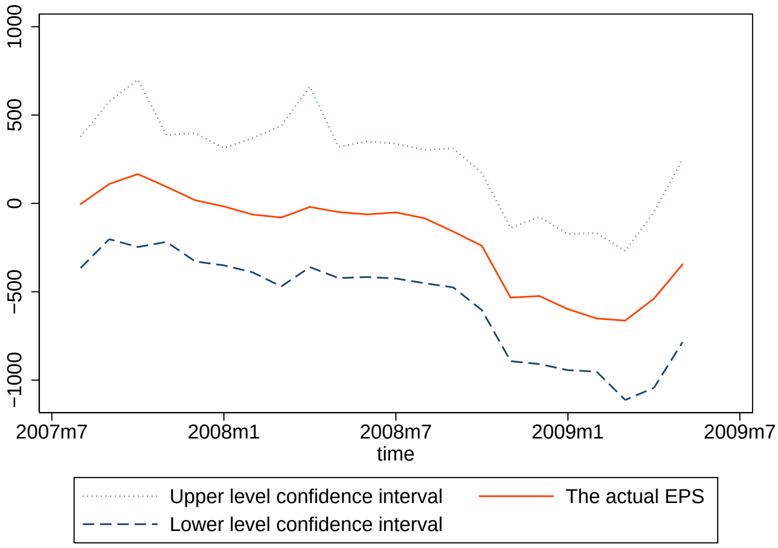
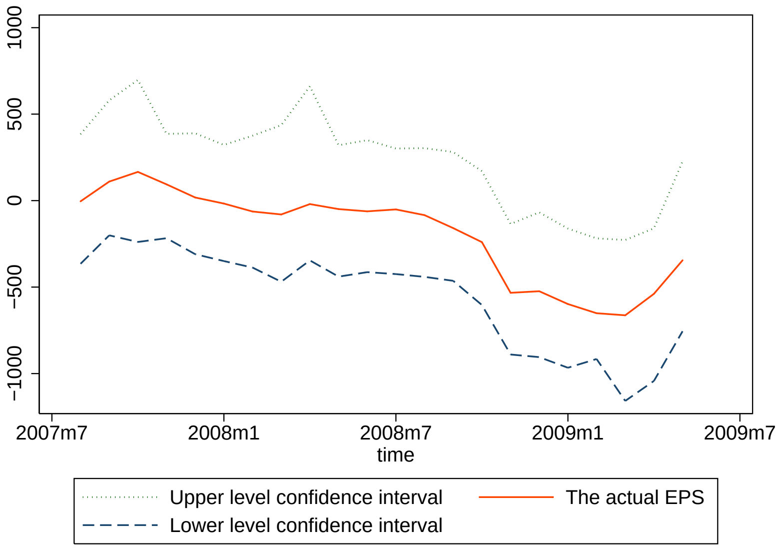


Figure 3.11: Bootstrap Confidence Intervals for Two-step-ahead Ordinary Forecast: 2007m8 — 2009m5



confidence intervals of the updated VAR while 22 actual values lie in the confidence intervals of the ordinary VAR. The results from bootstrap confidence intervals of one- and two-step-ahead forecasts suggest that confidence intervals for the ordinary VAR forecast are similar to those of the updated VAR forecasting.

To illustrate the asymptotic approximation of the predictor, we consider the same bivariate VAR model with the same coefficient parameters as a bootstrap approach. The main difference between a bootstrap and an asymptotic approach is that, for the former, it is necessary to estimate each bootstrap replaced by a bootstrap DGP from which to draw the simulated samples, whereas, for the latter, the corresponding critical value is chosen.

$$[fcast_l,\ fcast_u] = [f\hat{cast} - \sqrt{(1-\hat{\rho}_{12}^2)}\hat{s}_{fcast}\ c_\alpha,\ f\hat{cast} + \sqrt{(1-\hat{\rho}_{12}^2)}\hat{s}_{fcast}\ c_\alpha\ ],$$

where c_α is the asymptotic critical value. We choose $\alpha = 0.99$, so the corresponding critical value is 2.576. We repeat each simulation $R = 23$ times to calculate asymptotic confidence intervals for each forecasting period until we are through the whole dataset as well.

Figures 3.12 and 3.13 illustrate asymptotic confidence intervals of the updated VAR one-step-ahead forecasts and the ordinary VAR one-step-ahead forecasts respectively. Given 23 observations in the one-step-ahead forecasting time periods of 2007m7 to 2009m5, 22 actual values lie in the 99% asymptotic confidence intervals of the ordinary VAR while all 23 actual values lie in that of the updated VAR. Figures 3.14 and 3.15 demonstrate asymptotic confidence intervals of the updated VAR two-step-ahead forecasts and the ordinary VAR two-step-ahead forecasts respectively. Given 22 observations in the forecasting time periods of 2007m8 to 2009m5, 21 actual values lie in the 99% asymptotic confidence interval of the ordinary VAR while

Figure 3.12: Asymptotic Confidence Intervals for One-step-ahead Updated Forecast: 2007m7 — 2009m5

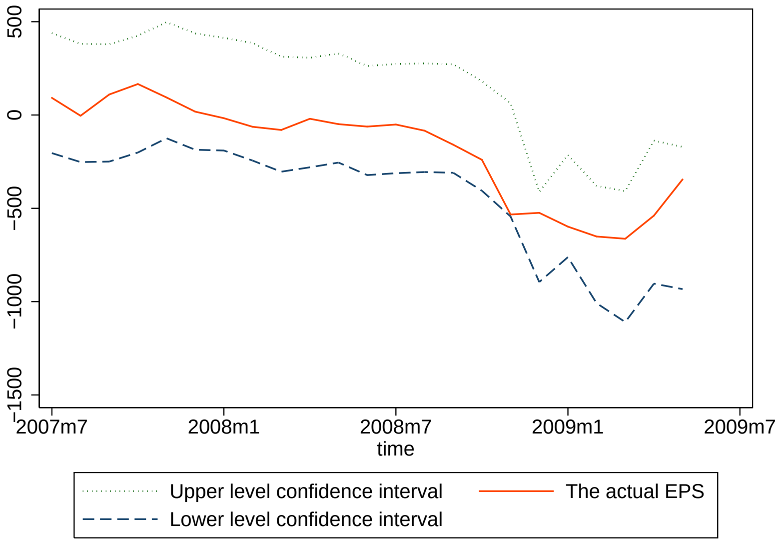


Figure 3.13: Asymptotic Confidence Intervals for One-step-ahead Ordinary Forecast: 2007m7 — 2009m5

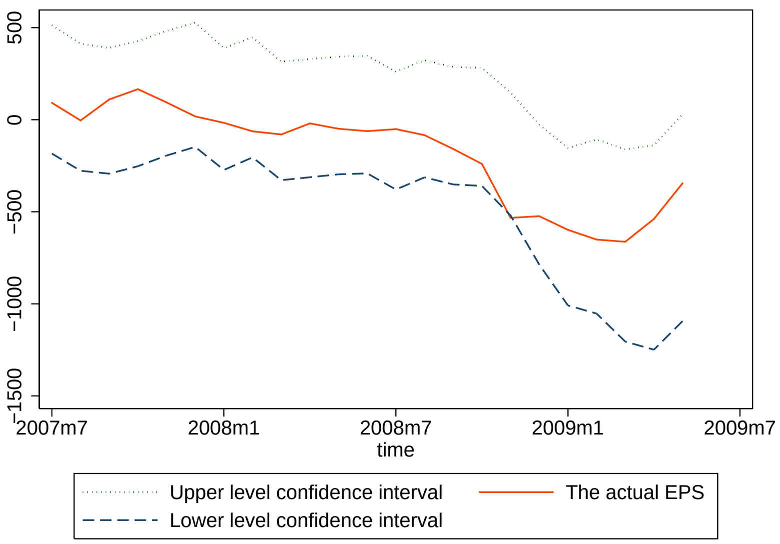


Figure 3.14: Asymptotic Confidence Intervals for Two-step-ahead Updated Forecast: 2007m8 — 2009m5

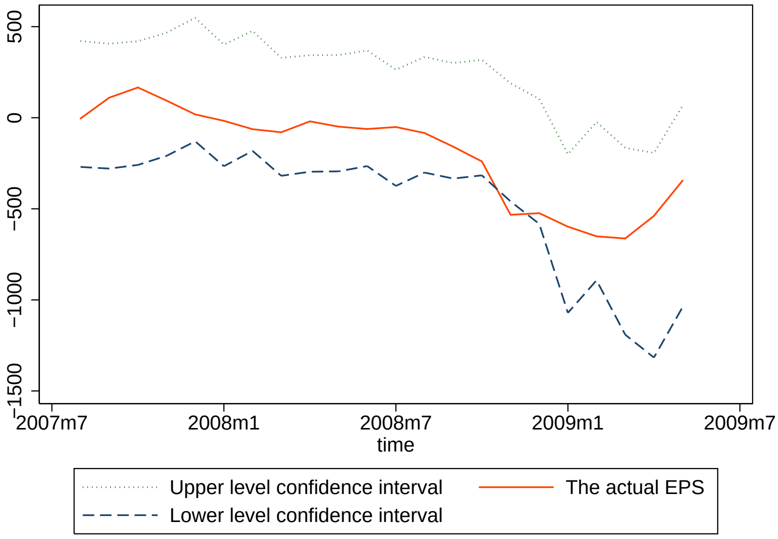
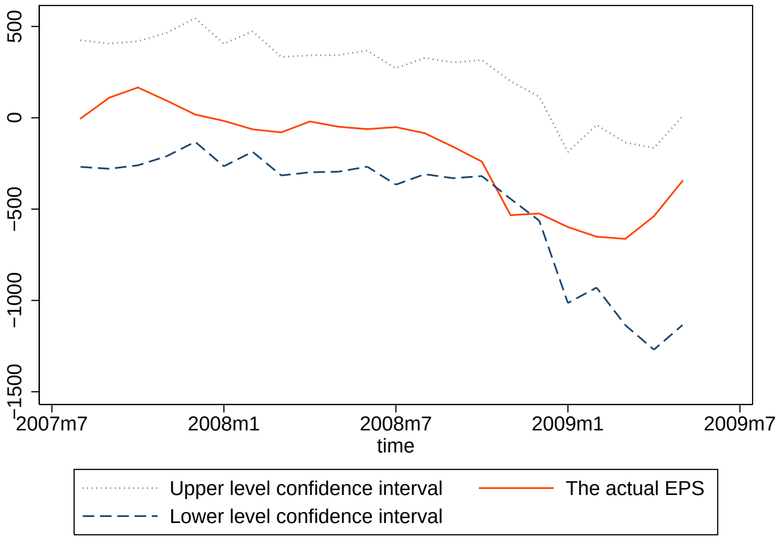


Figure 3.15: Asymptotic Confidence Intervals for Two-step-ahead Ordinary Forecast: 2007m8 — 2009m5



21 actual values lie in the 99% asymptotic confidence interval of the updated VAR. The results from asymptotic confidence intervals of one- and two-step-ahead forecasts indicate that confidence intervals for the updated VAR forecast is tighter than that of the ordinary VAR forecasting.

Overall the evidence suggests the confidence intervals are similar for both the asymptotic and bootstrap. Our guess is that for more complicated VARs with additional lags, the bootstrap procedure would likely dominate.

3.6 Monte Carlo Simulation Experiment

When the sample size is not large, hypothesis tests and confidence intervals based on asymptotic theory can be seriously misleading. In the previous example, there are certainly some suspicions that our sample sizes in the labor market are too small period. Often in these cases, confidence intervals tend to be too small even when the model is correctly specified, but asymptotic comparison has been used. Bootstrap methods are intended to overcome these shortfalls. In this section, we discuss a bootstrap approach and asymptotic confidence intervals in a simple experiment calculated against our labor example.

We consider the Monte Carlo experiment through the same bivariate model (3.1) with the same coefficient parameters, where $\hat{a}_{11} = 0.68, \hat{a}_{12} = 0.25, \hat{a}_{21} = 0.53$, and $\hat{a}_{22} = 0.41$. To observe how the correlation coefficient ρ affects the forecasting performance and confidence intervals in both the updated VAR and the ordinary VAR model, we choose $\rho = -0.7, -0.3, 0, 0.4$, and 0.8 respectively. We then generate data using the same procedure as in section 3.5 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 estimate the parameters and compute the forecasts using both the asymptotic and the bootstrap approach. We repeat each simulation $R = 1,000$ times for different correlations. Table 3.7 reports the MSE, average standard deviation, and the numbers of actual values falling in the intervals, such as 90%, 95% and 99% confidence intervals, for the bootstrap DGP. The values in the bracket indicate the number of actual values falling in the intervals based on the asymptotic approximation DGP.

Table 3.7: How Correlations Affect the Small Sample Behavior of the Predictor
Repetition $= 1,000$, *Bootstrap* $= 299$, and *Sample size* $= 100$

			% 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 1,000 times Monte Carlo simulation.

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 is extremely close to the actual value.

3.7 Conclusions

There are many applications in real-time forecasting. This chapter uses U.S. employment data to examine whether *adp* estimates, which are usually announced two days prior to *bls* estimates, are helpful in forecasting the total nonfarm payroll employment number in the same month by the BLS. Since data are frequently revised in this study, rather than use the final released data to estimate and forecast, we use data that are available when the real-time forecast is made. More specifically, at every date within a sample, variables in the model are the most up-to-date estimates at that time. Further, the first release values of BLS are what markets react to, so we adopt the notion of using the data at the time of forecasting to predict first release values of BLS. We find that the predicted employment number is more accurate in matching the labour data when considering the real-time information than the standard VAR. One reason for this is that the residuals of the VAR for BLS and ADP are highly correlated. As we have demonstrated theoretically, such a correlation can lead to a substantial fall in the MSE. Compared to the ordinary VAR quantitatively, the updated VAR forecast improves 8% on the one-step-ahead forecast and improves 16% on the two-step-ahead forecast. Simulation exercise reinforces the practical outcome and demonstrates that confidence intervals using either an asymptotic or bootstrap calculation to have appropriate coverage probability.

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.

Chapter 4

An Application to the Canadian Banking Industry

4.1 Introduction

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. According to the Office of the Superintendent of Financial Institutions, these institutions manage over \$2.6 trillion in assets. Table 4.1 shows that six banks dominate the market with 90.01% of all banking assets. Studying and forecasting these six banks' earnings is followed widely by investors interested in financial markets.

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. We focus on the six leading Canadian banks' EPS in order to study the Canadian banking industry, given the interest in these six banks, as well as their institutional different announcement dates for these earnings make this an excellent candidate for our methodology.

Table 4.1: The Big Six Canadian Banks Ranked by Asset Size

Ranking World* 2006	Ranking Canada** 2008	Name of Banks	Total Assets as of 31-Jan-08 $millions CDN	Percentage of Total Assets
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 banks ranked by asset size. Source: *The Banker*, July 2007
** Canadian bankers association. *Bank Financial Results: 2008 Fiscal Year*, April 2008

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.[1] 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

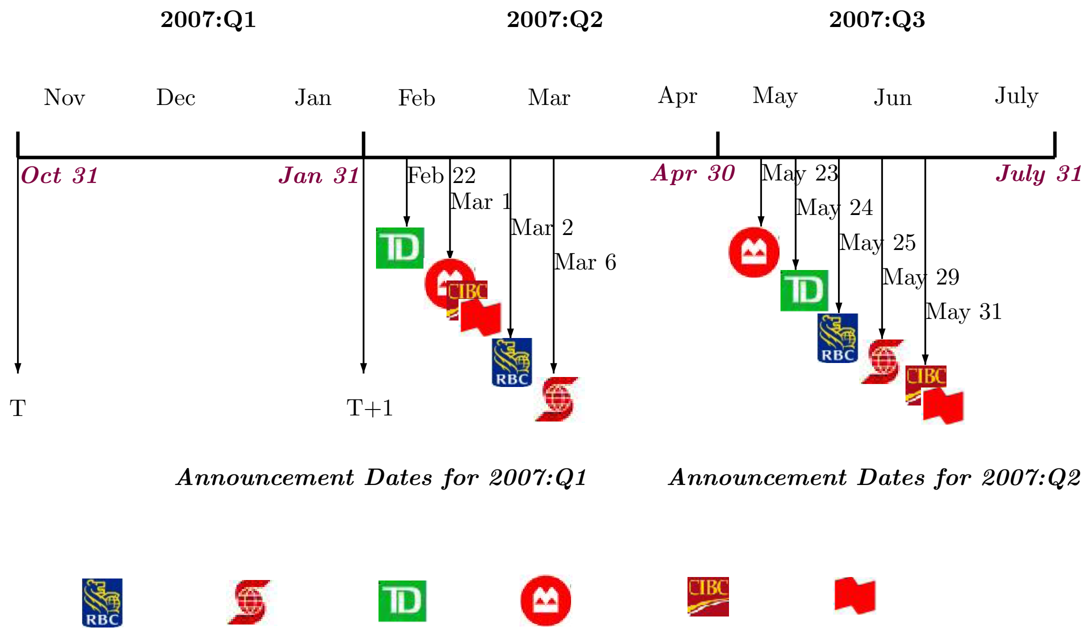
[1] http://en.wikipedia.org/wiki/Securities_Exchange_Act_of_1934.

for filing quarterly financial report statements of companies. Effective November 15, 2002, the filing deadline for quarterly reports was reduced gradually from 45 days to 35 days over three years under the Securities and Exchange Commission.[2] 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.[3] Therefore, each bank announces its periodic report within 35 days of the end of the quarterly period. Figure 4.1 illustrates the timeline of earnings release for the big six Canadian banks for the first and second quarter of 2007. For instance, at the first quarter of 2007, TD announced its first quarter earnings on Thursday, February 22 (which is a delay of 22 days from January 31), BMO, CIBC, and BNC announced their first quarter financial statements on Thursday, March 1 (which is a delay of 31 days after January 31), and RBC announced its first quarter earnings on Friday, March 2 (which is a delay of 32 days from January 31), and BNS announced its first quarter earnings on Tuesday, March 6 (which is a delay of 36 days from January 31). However, what we observe is that the sequential release of earnings for the second quarter of 2007 does not follow the same release order of the first quarter. It becomes BMO on May 23, TD on May 24, RBC on May 25, BNS on May 29, and CIBC and BNC on May 31. Thus the particular dates and ordering of banks' financial announcements vary. But as far as we can tell, announcement dates are exogenous to the earnings process itself (we examine in a limited way a departure from this assumption). This is the most important issue we deal with in the updated multivariate VAR forecasts.

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

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

Figure 4.1: Timeline of Earnings Release



4.2 Data

The big six Canadian banks' data set are collected from the I/B/E/S database through the 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 expressed in the first difference. We take the average of individual analysts' forecasts of EPS, and get the consensus forecast of EPS during 2007Q1 through 2009Q1 in comparison to the out-of-sample updated VAR forecast. Table 4.2 reports the descriptive statistics. Augmented Dickey-Fuller (ADF) nonstationarity tests have been conducted, and the presence of a unit root is rejected. All six series are stationary in first difference. Since 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 4.2. The plot of the quarterly changes of EPS is provided in Figure 4.3.

Figure 4.2 shows that banks' profits are correlated 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 had a significant impact on CIBC's financial performance for 2005. In the first quarter of 2008, there were two significant unexpected results. One is for CIBC whose profits drop dramatically.[4] The other unexpected

[4] "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

Table 4.2: 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 nonstationarity.

result is for BMO. Excluding significant items, net income of BMO is $617 millions. However, reported profits including significant items[5] are $255 millions.

4.3 Forecasting

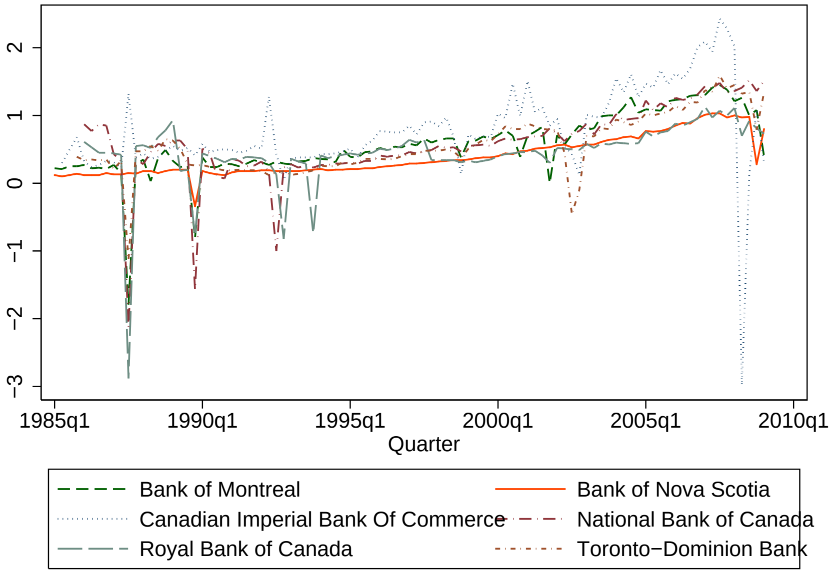
A VAR expresses each variable as a linear function of its own past values and the past values of all other variables being considered. Innovations can be contemporaneously correlated, which would suggest a VAR updating procedure could be useful. Thus, in this study the VAR involves six equations: *rbc*, *bns*, *td*, *bmo*, *cibc*, and *bnc*[6]. Each

value for our shareholders."

[5]Significant items are charges for certain trading activities and valuation adjustments and an increase in the general provision for credit losses.

[6]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.

Figure 4.2: Level of EPS

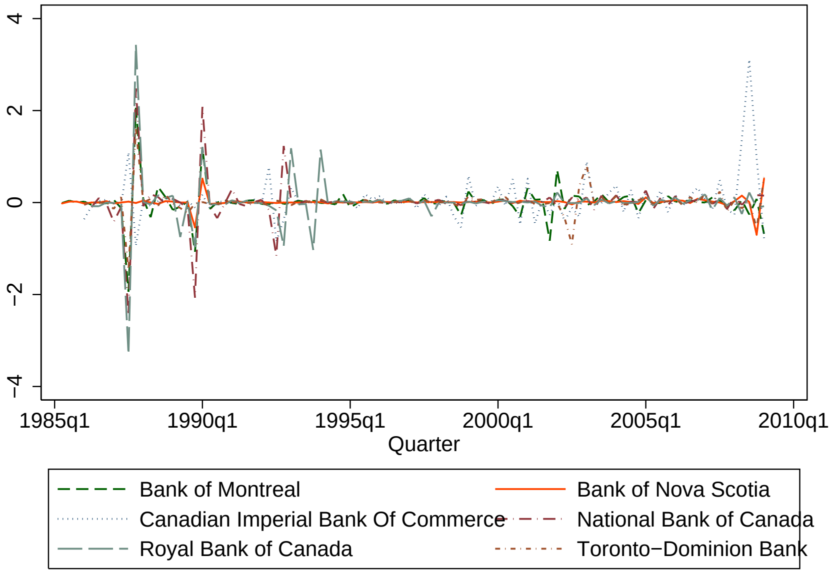


Source: I/B/E/S database through the Wharton Research Data Service and the current quarterly financial statements from each bank.

equation contains the same regressors and can be efficiently estimated by OLS. This OLS estimator is as efficient as the maximum likelihood estimator and the general least squares estimator. The number of lagged values to include in each equation is determined by Schwarz's Bayesian information criterion (SBIC). SBIC suggests that the optimal lag selection is one and we will use this throughout.

Each row in Table 4.3 represents an estimated VAR (1) equation. The first row demonstrates that the current earnings of RBC has a positive relationship with earnings of CIBC and BNC in the previous period and a negative relationship with earnings of itself, BNS, TD, and BMO in the last period. Except CIBC, the current earnings of the rest of the five banks all have a negative sign and a significant relationship (the bold number in Table 4.3) with the previous earnings of their own

Figure 4.3: Changes in EPS



banks. This indicates that earnings in the previous quarter have negative and significant explanation of the current quarter's earnings. The current quarter earnings of CIBC can be positively and significantly explained by the previous quarter's earnings of BNS and TD.

To facilitate the impact of the order of earnings release on earnings, we run a simple regression for earnings of each bank on its own orders of release. The regression results are shown in Table 4.4. It indicates that the release order is not random. In particular, the release orders of BNS, BMO, and BNC have a significant impact on their own earnings at 95% confidence interval level. This finding motivates us to study thought experiments in section 4.5

In the conventional VAR forecast, the coefficient estimates and the variance covariance matrix estimates are used to calculate the forecast mean squared error. In

Table 4.3: The Six Banks's EPS estimation

Estimates	rbc_{t-1}	bns_{t-1}	td_{t-1}	bmo_{t-1}	$cibc_{t-1}$	bnc_{t-1}
rbc_t	**-0.55**	-0.21	-0.02	-0.11	+0.06	+0.22
	(0.18)	(0.53)	(0.28)	(0.29)	(0.12)	(0.19)
bns_t	-0.05	**-0.63**	+0.01	+0.04	-0.10	+0.02
	(0.03)	(0.10)	(0.05)	(0.05)	(0.02)	(0.03)
td_t	-0.14	-0.19	**-0.24**	+0.12	-0.02	-0.02
	(0.09)	(0.27)	(0.14)	(0.15)	(0.06)	(0.10)
bmo_t	+0.01	+0.38	+0.10	**-0.50**	-0.05	-0.07
	(0.12)	(0.35)	(0.18)	(0.19)	(0.08)	(0.13)
$cibc_t$	+0.02	+0.84	+0.52	-0.38	+0.13	+0.09
	(0.16)	(0.47)	(0.24)	(0.25)	(0.10)	(0.17)
bnc_t	-0.15	-0.33	+0.07	+0.04	-0.03	**-0.37**
	(0.16)	(0.48)	(0.25)	(0.26)	(0.11)	(0.17)

† The entries without brackets are coefficient estimates.
† The entries with brackets are the standard error of coefficient estimates.

the updated VAR forecast, in addition to coefficient estimates and variance covariance matrix estimates, we also consider the correlation of residuals of the VAR to compute the more efficient forecast mean squared error. The correlation of residuals of the multivariate VAR is shown in Table 4.5, indicating a positive correlation between all banks, except that CIBC has negative correlations with three of the other banks.

The big six banks' EPS are covariance stationary if their first two moments exist and are independent of time. Table 4.6 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 time series follow a VAR(1) stationary process.

Further, a test for normally distributed disturbances after fitting VAR (1) model is conducted, for each equation and for all equations jointly. Table 4.7 reports a series of statistics — a skewness statistic, a kurtosis statistic, and the Jarque-Bera statistic —

Table 4.4: Impact of the Release Order on Earnings

	constant	*index*
RBC	0.68	-0.07
	(0.13)	(0.04)
BNS	0.17	0.08*
	(0.07)	(0.02)
TD	0.52	0.02
	(0.10)	(0.06)
BMO	1.10	-0.16*
	(0.08)	(0.02)
CIBC	0.85	-0.02
	(0.20)	(0.05)
BNC	0.29	0.11*
	(0.12)	(0.04)

Standard errors are given in the parentheses. * represents statistical significance at the 95% confidence interval level. The null hypothesis, the coefficient estimate of "*index*" variable is zero, is rejected and it is concluded that the release orders of BNS, BMO, and BNC have a significant impact on their own earnings.

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 innovations are multivariate normal.

Since all the banks announce their quarterly earnings at different dates and conform to our key assumptions, the theoretical updating VAR framework in section 2 is an approach appropriate 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

Table 4.5: Correlations of Residuals

	RBC	BNS	TD	BMO	CIBC	BNC
Correlation Matrix						
RBC	1.00					
BNS	0.22	1.00				
TD	0.67	0.22	1.00			
BMO	0.84	0.26	0.58	1.00		
CIBC	-0.20	0.05	-0.18	-0.22	1.00	
BNC	0.76	0.39	0.54	0.79	-0.10	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	-0.05	0.01	-0.02	-0.03	0.02	
BNC	0.18	0.02	0.06	0.12	-0.02	0.21

one quarter ahead; *i.e.*, forecasts 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. We calculate forecast mean square errors and compare our findings to the ordinary VAR and the consensus forecast. The consensus forecast is a simple arithmetic average of all of the individual predictions collected by I/E/B/S.

The detailed estimation method is investigated in the following two-stage one-step-ahead VAR forecast. In the first stage, we estimate the VAR over the period from 1986Q2 through 2006Q4. The fact that the different banks release dates at different points requires the updating method to estimate various accessible regressors. The method follows section 2.1.3 depending on the number of releases (the order) and the number of banks to forecast under the updates. In cases like multiple releases, we

Table 4.6: 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 the stability condition.

produce a sequence of updated forecasts for some banks.

We consider forecast periods from 2007Q1 through 2009Q1, which yields nine observations. Table 4.8 shows the actual order of earnings release during the forecasting time period. For instance, 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 rest of the data

Table 4.7: 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

for release is available in Table 4.8.

Table 4.8: Actual Sequential Release, Forecasting time period: 2007Q1 — 2009Q1

	RBC	BNS	TD	BMO	CIBC	BNC
07:Q1	3	4	1	2	2	2
07:Q2	3	4	2	1	5	5
07:Q3	2	3	1	3	4	4
07:Q4	3	4	2	1	4	2
08:Q1	2	3	1	3	1	1
08:Q2	3	1	2	1	3	3
08:Q3	3	1	3	1	2	3
08:Q4	4	2	3	1	3	3
09:Q1	2	3	1	3	2	2

† The entries are the actual order of earnings release during the forecasting time period from 2007Q1 through 2009Q1. For instance, the order of earnings release on 2007Q1 is TD first, followed by BMO, BNC, and CIBC simultaneously, then RBC, and finally BNS.

Figures 4.4 to 4.9 show earnings forecasts of the big six Canadian banks (RBC, BNS, TD, BMO, CIBC, and BNC respectively) from the first quarter of 2007 to the first quarter of 2009. In order to illustrate forecasting results, we investigate the updating earnings forecasts, the conventional earnings forecasts, and the actual earnings from time period 2007Q1 through 2009Q1 in each figure. Figure 4.4 shows that the updating earnings forecasts of RBC are close to the actual earnings from the forecasting period of 2007Q1 through 2008Q4. Although the updating forecast could not fit the actual earnings on 2009Q1, it is still better than the conventional forecast. The overall updating earnings forecast root mean squared error (FRMSE) of RBC is 0.39, which has a 26% relative efficiency gain in comparison to the conventional VAR forecast.

Figure 4.5 illustrates the updating earnings forecasts, the conventional earnings forecasts, and the actual earnings of BNS from the forecasting period of 2007Q1

Figure 4.4: Earnings Forecasts of the Canadian Banks — RBC

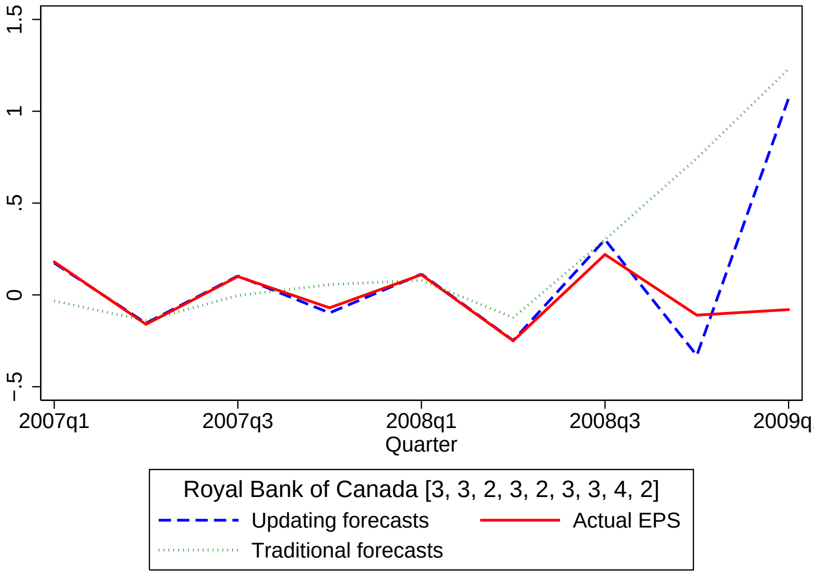


Figure 4.5: Earnings Forecasts of the Canadian Banks — BNS

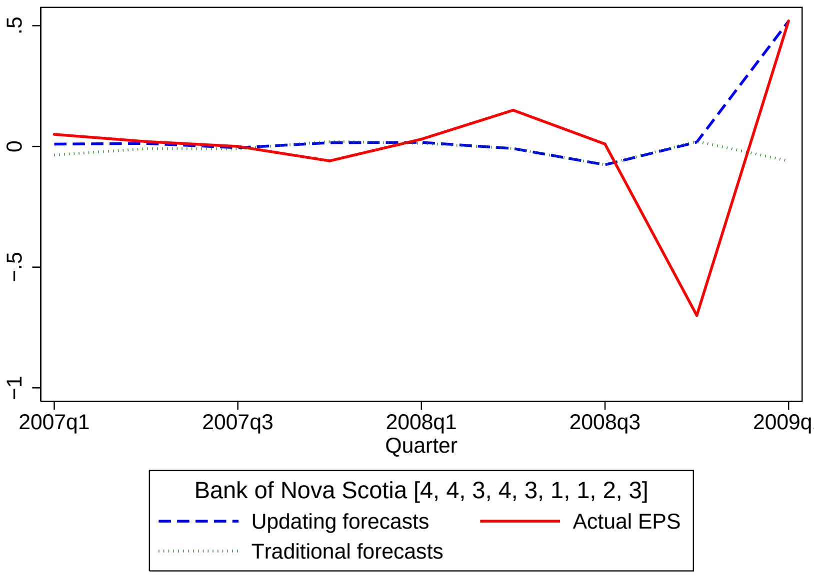


Figure 4.6: Earnings Forecasts of the Canadian Banks — TD

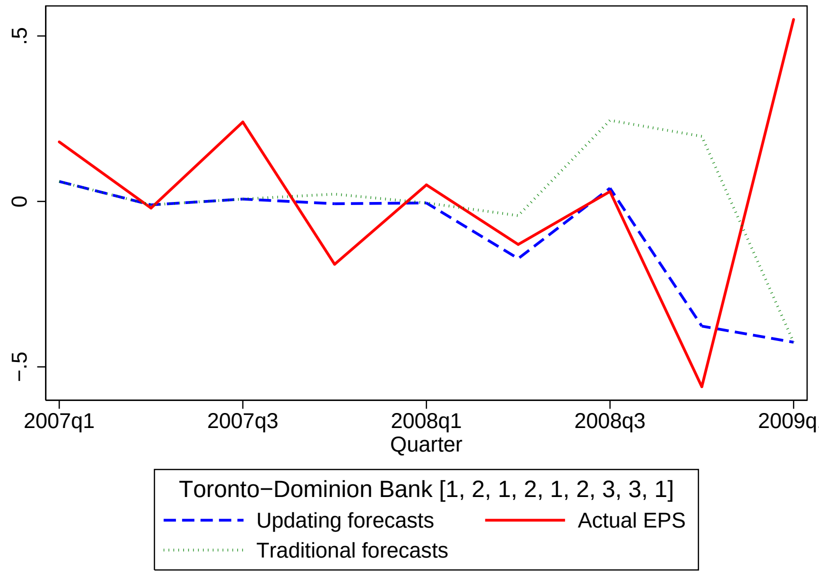


Figure 4.7: Earnings Forecasts of the Canadian Banks — BMO

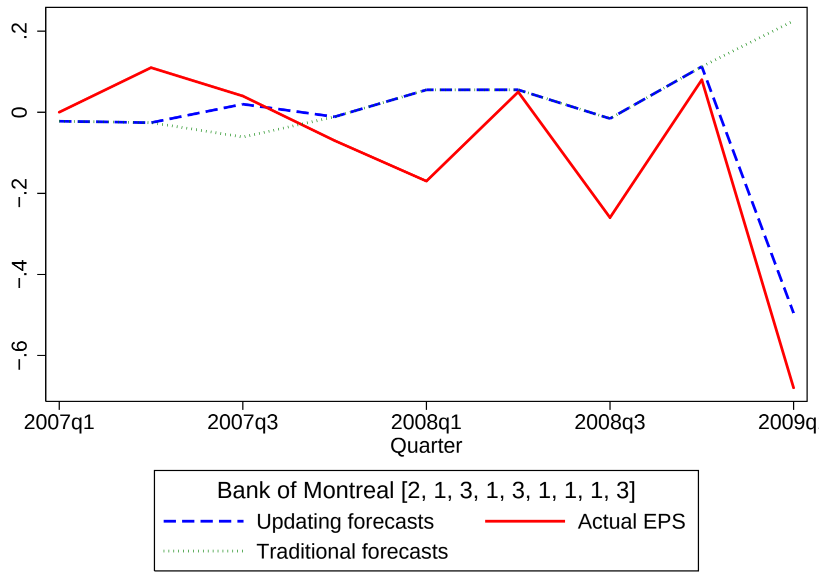


Figure 4.8: Earnings Forecasts of the Canadian Banks — CIBC

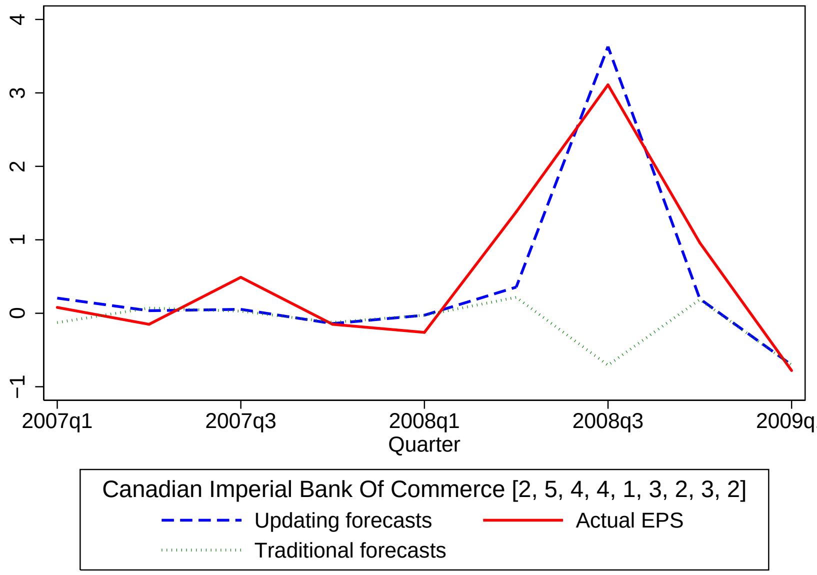
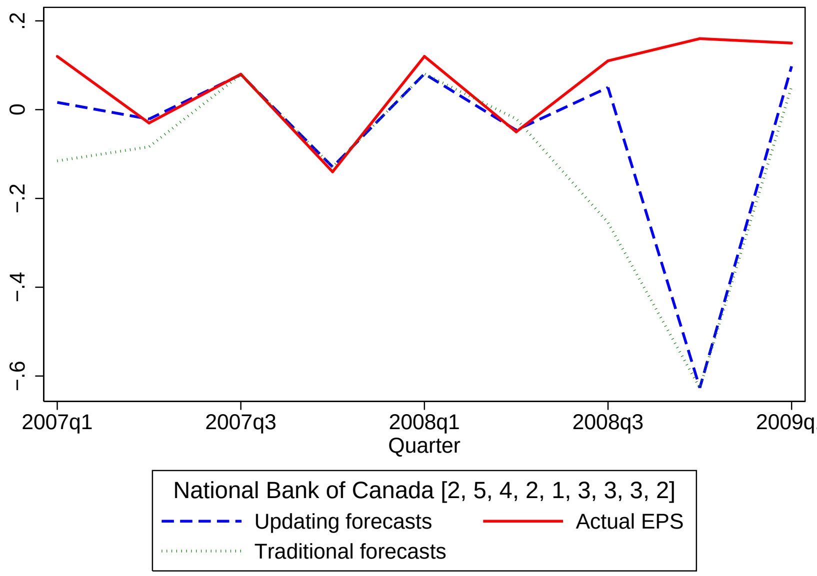


Figure 4.9: Earnings Forecasts of the Canadian Banks — BNC



through 2009Q1. We find that the updating VAR forecasts are close to the actual earnings in the most recent forecasting time period except on 2008Q4. The innovation of BMO which released its announcement on 2008Q4 prior to BNS does not provide useful information about the innovation of BNS, but the updating forecast at this period is at least as good as the conventional forecast. The overall updating earnings FRMSE of BNS is 0.25, which represents a 22% relative efficiency gain compared to the conventional forecast.

Figure 4.6 demonstrates the earnings forecasts of TD. The updating earnings forecasts again are close to the actual earnings except on 2009Q1. Since TD first released its earnings on the first quarter of 2009, the advantage of sequential release information no longer exists. The updating earnings FRMSE of TD is 0.35, which is a 19% relative efficiency gain in comparison to the conventional forecast. The updating earnings forecast of BMO perform well through the whole forecasting time period. In particular, the updating innovation provided useful information on 2009Q1. Hence, the updating forecast moves in the same direction as the actual earnings, shown in Figure 4.7. The relative efficiency gain of the updating earnings forecast of BMO is 58%, compared to the conventional forecast. Again, we make better forecasts for CIBC compared to the conventional forecast, especially on 2008Q3. BMO released its earnings on 2008Q3 prior to CIBC and provided useful information about innovation. Thus, the updating earnings forecast catch up in the same direction as the actual earnings. As a consequence, the updating FRMSE is 0.49 with a 64% relative efficiency gain compared to the conventional forecast.

Figure 4.9 indicates that on 2008Q4, we could not learn useful information about innovations for BNC prior to announcements by BMO and BNS, who released their earnings on the first and second order of sequential release respectively. The updating earnings forecasts capture the most of the actual earnings during the forecasting time period. The relative efficiency gain of the updating earnings forecast of BNC is 10% in comparison to the conventional forecast from forecasting time period 2007Q1 through 2009Q1. The third column of Table 4.9 shows the relative efficiency gain of the updating earnings forecasts compared to the conventional forecasts.

4.4 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 4.9 reports the root mean squared errors for each six banks by various methods: the updated VAR, the ordinary VAR, consensus forecasts, and contemporaneous regression forecasts. The third column represents the efficiency gains of the updated VAR relative to the ordinary VAR, whereas the fifth column demonstrates the efficiency gains of the updated VAR compared with consensus forecasts. The seventh column illustrates the efficiency gains of the updated VAR compared with contemporaneous forecasts. We find that the RMSEs for each bank using the updated VAR forecast are smaller than those using the ordinary VAR forecast and the contemporaneous regression forecast. The maximum efficiency gain of the six banks is 64% by CIBC. Although CIBC has a 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. By using the updated VAR forecast, we find that the RMSEs of RBC, BNS, BMO, and CIBC are typically smaller than the RMSEs of consensus forecasts respectively. There are typically certain banks (TD and BNC) where the consensus forecast is preferred.

The contemporaneous regression model uses current earnings of firms' data that has been released to predict current earnings of one firm that has not been released. The theoretical findings in section 2.2 show that the contemporaneous model is misspecified, and hence is not a good forecasting strategy. In comparison to contemporaneous regression forecasts, the updated VAR all have smaller RMSEs for each five individual banks. The relative efficiency gains are 26%, 26%, 39%, 67%, 65%, and 56% for RNC, BNS, TD, BMO, CIBC, and BNC respectively. This is consistent with the theoretical findings in section 2.2.

Table 4.9: Forecast Root Mean Squared Error

	M(1)	M(2)	$Gains^1$	M(3)	$Gains^2$	M(4)	$Gains^3$
RBC	0.53	0.39	26%	0.40	3%	0.53	26%
BNS	0.32	0.25	22%	0.29	14%	0.33	26%
TD	0.43	0.35	19%	0.28	-20%	0.57	39%
BMO	0.33	0.14	58%	0.26	46%	0.42	67%
CIBC	1.37	0.49	64%	1.16	58%	1.40	65%
BNC	0.30	0.27	10%	0.12	-56%	0.61	56%

†M(1) refers to the traditional VAR forecast; M(2) represents the updating VAR forecast; M(3) is the professional consensus forecast; M(4) refers to the contemporaneous regression forecast.
† $Gains^1$ is the relative efficiency gain of the updating forecast to the traditional forecast; $Gains^2$ is the relative efficiency gain of the consensus forecast to the traditional forecast; $Gains^3$ is the relative efficiency gain of the contemporaneous regression forecast to the traditional forecast.

4.5 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 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 maximum disclosure.

We consider the forecast period of 2007Q1 through 2008Q1, consisting of five one-step-ahead forecast observations. The first column in Table 4.10 is the RMSEs using the updated VAR forecast with the true release order. The rest of five columns are the RMSEs using the updated VAR forecast with a different ordering for each bank which lead to the lowest RMSE. 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 that produces the smallest MSE, 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 again with the smallest MSE. Interestingly most of the gain in accuracy comes from the release of a bank with innovations that are highly correlated with the one forecasted; other subsequent releases generally do not provide much more accuracy. 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 of 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 of banks by using the updated VAR forecast method.

Table 4.10: Reordering of Announcement Interval

		Updating VAR				
	Actual Release	$RMSE^1$	$RMSE^2$	$RMSE^3$	$RMSE^4$	$RMSE^5$
RBC	0.39	0.03	0.02	**0.01**	0.03	0.06
BNS	0.25	0.04	0.03	0.03	**0.02**	0.03
TD	0.35	0.11	0.10	**0.09**	0.10	0.16
BMO	0.14	**0.10**	0.11	0.11	0.12	0.14
CIBC	0.49	0.26	0.26	0.24	0.26	**0.23**
BNC	0.27	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 column $RMSE$ are the updating forecast root mean squared errors based on the true release order. The entries in the columns 2, 3, 4, 5, and 6 respectively are the forecast root mean squared errors (RMSE) for 1, 2, 3, 4, 5 release for the bank(s) that lead to the lowest RMSE.

4.6 Does the Order Matter

The order of banks' earnings release is known and exogenous by assumption, but we can investigate the impact of the ordering in a limited way and still retain the crucial aspects of the updating method. In this section, we include a value that indexes the order in which the forecast is released. The variable would take on a value in each regression for the order in which it releases for as far as we know at the beginning of 1986. The VAR model becomes

$$Y_t = C + A \times Y_{t-1} + B \times INDEX_t + \epsilon_t, \qquad t = 1, 2, ..., T, \tag{4.1}$$

where $Y_t = (rbc_t, bns_t, td_t, bmo_t, cibc_t, bnc_t)'$, $INDEX_t = (index_{rbc,t}, index_{bns,t}, indext_{td,t}, index_{bmo,t}, index_{cibc,t}, index_{bnc,t})$, A is (6×6) coefficient estimates, and B is a (1×6) random vector. $INDEX_t$ corresponds to the order of earnings release for one of the six banks while the coefficient estimates B of orders of earnings release correspond to the sign and magnitude of impart on earnings. For data outside the known order the index is set to 1. C is an intercept.

Including a value that indexes the order in which the forecast is released changes both the updated forecast RMSE and the conventional forecast RMSE. Table 4.11 shows that in the conventional VAR forecast, the RMSEs of TD and BNC with indexing the order as exogenous variables are slightly larger than the RMSEs of TD and BNC without indexing the order respectively. The RMSEs of the rest of four banks with indexing the order are slightly smaller than the RMSEs of the corresponding banks. In the updated VAR forecast, the RMSE of RBC with indexing the order is slightly smaller than that of RBC without indexing the order. The RMSEs of the rest of five banks with indexing the order are larger than those without indexing the order. For the most part, we find that the indexing variable does not help forecasting and RMSE may rise purely because of events that are unrelated to order.

Table 4.11: The Order Matters with Sequential Release of Information

	Updating VAR		Ordinary VAR	
	w/o the order	w/ the order	w/o the order	w/ the order
RBC	0.39	0.36	0.53	0.49
BNS	0.25	0.51	0.32	0.31
TD	0.35	0.38	0.43	0.44
BMO	0.14	0.15	0.33	0.31
CIBC	0.49	1.26	1.37	1.33
BNC	0.27	0.28	0.30	0.31

4.7 Optimal Ordering of Announcement

To address an optimal ordering of announcement for firms, we will use the entire VAR sample period to estimate all magnitudes. Each combination of real-time information is required. In general, with N firms there will be $N!$ combinations of the real-time ordering of announcements. With six banks there are 720 updates. It is too complicated to examine optimal ordering over all possibilities. We consider $N = 3$ and calculate the 3 banks to examine numerically results in Chapter 2 (Propositions 6 and 7).

In this section, we examine the finite sample performance of the release sequence using Monte Carlo simulations. We choose the parameters which are obtained by estimating three Canadian banks' quarterly earnings per share data in the sample periods of 1986Q2 to 2008Q1. These three Canadian banks are Royal Bank of Canada, Bank of Nova Scotia, and Toronto-Dominion Bank. We adopt simulation designs that satisfy the following conditions.

$$\begin{pmatrix} \triangle RBC_t \\ \triangle BNS_t \\ \triangle TD_t \end{pmatrix} = \begin{pmatrix} -0.46^* & -0.18 & -0.05 \\ -0.02 & -0.47^* & 0.06 \\ -0.13 & 0.24 & -0.14 \end{pmatrix} \begin{pmatrix} \triangle RBC_{t-1} \\ \triangle BNS_{t-1} \\ \triangle TD_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{RBC,t} \\ \epsilon_{BNS,t} \\ \epsilon_{TD,t} \end{pmatrix}$$

The residual correlation coefficients are $\rho_{rbc,bns} = 0.27$, $\rho_{rbc,td} = 0.69$, and $\rho_{bns,td} = 0.08$. The residual variances are $\sigma^2_{rbc} = 0.27$, $\sigma^2_{bns} = 0.07$, and $\sigma^2_{td} = 0.01$.

We consider two cases. The first simple case is to set $\sigma_{rbc} = \sigma_{bns} = \sigma_{td} = 1$. The second general case is to set $\sigma^2_{rbc} > \sigma^2_{bns} > \sigma^2_{td}$, where we obtain the residual variances from three Canadian banks. To fit our setting, we set $\sigma^2_{rbc} = 0.27$, $\sigma^2_{bns} = 0.07$, and $\sigma^2_{td} = 0.01$. We generate data from the following model with sample size $T+1 = 1,001$ for both cases, where the value at time $T+1$ is taken as either the out-of-sample real

Table 4.12: Optimal Ordering 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 $\sigma_{rbc} = \sigma_{bns} = \sigma_{td} = 1$. The third column 'Case II' is the case where the residual variances are $\sigma^2_{rbc} = 0.27$, $\sigma^2_{bns} = 0.07$, and $\sigma^2_{td} = 0.01$.
† S.D. denotes the standard deviation.

value for comparison or real-time information we know in advance.

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

We run 10,000 experiments. The simulation results are reported in Table 4.12.

We set the assumptions in 'Case I' the same as Proposition 6, where $\sigma_{rbc} = \sigma_{bns} = \sigma_{td} = 1$ and $\rho_{rbc,td} > \rho_{rbc,\ bns} > \rho_{bns,\ td}$. We find that there is a 32% possibility that the disclosure sequence of RBC, BNS, and TD will have the smallest sum of forecast squared errors. We see that sequence of $\{rbc,\ bns,\ td\}$ has the largest probability of having the minimum sum of forecast squared errors. The result induces the conclusion that correlation coefficients do play a role in determining the maximum disclosure order.

Given the assumptions $\sigma^2_{rbc} > \sigma^2_{bns} > \sigma^2_{td}$ as well as $\rho_{rbc,td} > \rho_{rbc,bns}$ and $\rho_{rbc,td} > \rho_{bns,td}$ in 'Case II', the simulation results show that the disclosure sequence, RBC, TD, and BNS, has the minimum sum of forecast squared errors with 100% probability since the variance of 1 is so large. In other words, based on the information release criterion — minimize standardized forecast squared errors at each informational release — the maximum disclosure order is unique (RBC, followed by TD, and then BNS). These simulation results coincide with our theoretical finding in Proposition 7.

4.8 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.

This updated multivariate 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 chapter examines whether one bank's earnings announced at the current quarter 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 chapter 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 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 the 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.

Chapter 5

Conclusions and Extensions

The easiest way to understand this area of research is to consider how investors, economists and planners adjust their outlooks with the arrival of new financial data. To be concrete and to follow the specific application addressed in the thesis, consider the earnings announcements of the six largest Canadian banks. The quarterly financial statements of the six Canadian banks are supervised by the Exchange Act. Under the U.S. Security and Exchange Commission, banks are mandated to announce their earnings within 35 days after the end each quarter. That is, banks' earnings for the same period (the previous quarter) are announced over an interval of time (usually a couple of weeks in Canada). The first, second, third, and fourth quarter financial results for the Canadian banking industry end on January 31, April 30, July 31, and October 31, respectively. On February 22 2007, TD first announced its earnings for 2007Q1. The critical forecasting question is: how can we best incorporate the information from the TD earnings release to revise our forecasts of earnings for 2007Q1 for the remaining 5 banks?

One might expect that if the TD earnings were, say, below expectation or street consensus that market observers would lower their earnings expectations for the remaining banks. This thought process of "adjusting" earnings in light of new information is what our research formally accomplishes. In fact, one key element of this research is to isolate the "surprise" element of an announcement and to determine how much to adjust the forecasts of the remaining banks. This process of revising the forecast can be done sequentially as further banks release their earnings until all banks' earnings for the information cycle are complete. Then the next quarter brings a new cycle.

Earnings are but one example of the kind of application that can be considered in this very general framework. For example, in Chapter 3 of the thesis, we consider forecasting the monthly nonfarm payroll employment from the Bureau of Labor Statistics (BLS) which comes out the first Friday of every month. This is done by using monthly employment for private firms based on surveys from Automatic Data Processing, Inc. (ADP) which are announced on the first (preceding) Wednesday of each month. Unlike the previous bank examples, announcement dates are fixed. This two-day prediction window is used to revise the forecast of the BLS data.

One ambitious area of future research that fits into this general setting is the time separation of markets on a daily basis. We know that markets in Europe are open in the day before those in United States and market observers often wish to revise their views or outlooks for the daily U.S. market in light of what has happened earlier in that day (including those in Asia). This framework can incorporate such daily revisions in a systematic way and highlight the nature of the "new" information hitting the market on a daily basis. We think that this is an exciting and wide-open application for this method.

Bibliography

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.

Atiase, R.K., H. Li, S. Supattarakul and S. Tse, 2005, "Market Reaction to Multiple Contemporaneous Earnings Signals: Earnings Announcements and Future Earnings Guidance", *Review of Accounting Studies*, 10, 4, 497-525.

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.

Boyd, J. H., J. Hu and R. Jagannathan, 2005, "The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks", *Journal of Finance* LX, 649-672.

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.

Chen, S., D.A. Matsumoto, and S. Rajgopal, 2006, "Is Silence Golden? An Empirical Analysis of Firms that Stop Giving Quarterly Earnings Guidance", Available at SSRN: http://ssrn.com/abstract=820644.

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.

Faust, J. et al., 2007 "The high-frequency response of exchange rates and interest rates to macroeconomic announcements", *Journal of Monetary Economics* 54, 1051-1068.

Flannery, M. J. and A. A. Protopapadakis, 2002, "Macroeconomic factors do influence aggregate stock returns", *Review of Financial Studies* 15, 751-782.

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.

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.

Kitagawa, G. and W. Gersch, 1996, *Smoothness Priors Analysis of Time Series*. Springer: New York.

Krueger, A. B., 1996, "Do markets respond more to more reliable labor market data? A test of market rationality", NBER Working paper 5769.

Lee, Charles M. C., J. Myers, and B. Swaminathan, 1999, "What is the intrinsic value of the Dow?" *Journal of Finance* 54 (5), 1693-1741.

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.

Milian, J.A., 2008, "The Relative Information Content of Guidance and Earnings", Available at SSRN: http://ssrn.com/abstract=1274103.

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.

Rapach, D. E., M. E. Wohar, J. Rangvid, 2005, "Macro variables and international stock return predictability", *International Journal of Forecasting* 21, 137-166.

Sarno, L. and M. Taylor, 1998, "Real exchange rates under the recent float: unequivocal evidence of mean reversion", *Economics Letters* 60, 131-137.

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.

————, 1989, "New indexes of coincident and leading economic indicators", in J. Olivier and Stanley Fisher, eds.: NBER Macroeconomics Annual 1989, Cambridge, Massachusetts.

Triantafyllopoulos, K., 2007, "Covariance estimation for multivariate conditionally Gaussian dynamic linear models", *Journal of Forecasting* 26, 551-569.

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.

Appendix A

Proofs for Chapter 2

Proof of proposition 1. Given the information set $\{y_{11}, y_{12}, ..., y_{1T+1}, y_{21}, ..., y_{2T}\}$, 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 information set $\{y_{11}, y_{12}, ..., y_{1T+1}, y_{21}, ..., y_{2T}\}$, 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 (2.1). We can obtain the moving average model notation by iterating forward.

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

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

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

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

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

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 (A.1) 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

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

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

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

■

Proof of proposition 3. Given the information set $\{y_{11}, y_{12}, ..., y_{1T+1}, ..., y_{1T+s}, y_{21}, ..., y_{2T}\}$, 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 (2.7) 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

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

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 (2.7) 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

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

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

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

■

Proof of proposition 4. The information set is $\{y_{11}, ..., y_{1T}, y_{1T+1}; ...; y_{m1}, ..., y_{mT}, y_{mT+1}; y_{m+1,1}, ..., y_{m+1,T}; ...; y_{N1}, ..., y_{NT}\}$. Rearrange equation (2.15) at time $T+1$, the following equation holds.

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

Given the known information set $\{y_{11}, ..., y_{1T}, y_{1T+1}; ...; y_{m1}, ..., y_{mT}, y_{mT+1}; y_{m+1,1}, ..., y_{m+1,T}; ...; y_{N1}, ..., y_{NT}\}$, we know ϵ_{mT+1}. All the assumptions hold, The multivariate prediction equation becomes

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

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

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

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

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

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,[1] the conditional variance ϵ_j given ϵ_i is the Schur complement of $\Omega_{\epsilon_i \epsilon_i}$ in Ω. That is,

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

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

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

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

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

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

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

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

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

■

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 value of variable, that is, $MSE[y^{1:2}_{4t}(1)] < MSE[y^{1}_{4t}(1)]$. By equation

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

(2.10), the forecast MSE with knowing one value of variable is

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

By equation (2.13), the forecast MSE with knowing the value of two variables is

$$MSE[y_{4t}^{1:2}(1)] = (1 - \underbrace{\frac{\rho_{14}^2 + \rho_{24}^2 - 2\rho_{14}\rho_{24}\rho_{12}}{1 - \rho_{12}^2}}_{E})\sigma_4^2. \tag{A.3}$$

To show equation (A.2) is larger than equation (A.3), 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

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

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 (2.16),

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

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

$$\Omega_{\epsilon_4\epsilon_3} = \begin{pmatrix} \rho_{41}\sigma_4\sigma_1 & \rho_{42}\sigma_4\sigma_2 & \rho_{43}\sigma_4\sigma_3 \end{pmatrix}.$$

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

$$\epsilon_{iT+1} = \begin{pmatrix} \epsilon_{1T+1} \\ \epsilon_{2T+1} \\ \epsilon_{3T+1} \end{pmatrix}.$$

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

$$\begin{aligned} \Omega_{\epsilon_3\epsilon_3} &= \begin{pmatrix} Var(\epsilon_1) & Cov(\epsilon_1, \epsilon_2) & Cov(\epsilon_1, \epsilon_3) \\ Cov(\epsilon_2, \epsilon_1) & Var(\epsilon_2) & Cov(\epsilon_2, \epsilon_3) \\ Cov(\epsilon_3, \epsilon_1) & Cov(\epsilon_3, \epsilon_2) & Var(\epsilon_3) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}. \end{aligned}$$

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

$$\Omega^{-1}_{\epsilon_3\epsilon_3} = \frac{1}{(1-\rho_{23}^2+\rho_{12}(\rho_{23}\rho_{31}-\rho_{21})+\rho_{13}(\rho_{12}\rho_{23}-\rho_{13}))\sigma_1^2\sigma_2^2\sigma_3^2} \times \begin{pmatrix} (1-\rho_{23}^2)\sigma_2^2\sigma_3^2 & (\rho_{23}\rho_{31}-\rho_{12})\sigma_1\sigma_2\sigma_3^2 & (\rho_{12}\rho_{23}-\rho_{13})\sigma_1\sigma_2^2\sigma_3 \\ (\rho_{23}\rho_{31}-\rho_{12})\sigma_1\sigma_2\sigma_3^2 & (1-\rho_{13}^2)\sigma_1^2\sigma_3^2 & (\rho_{13}\rho_{21}-\rho_{23})\sigma_1^2\sigma_2\sigma_3 \\ (\rho_{12}\rho_{23}-\rho_{13})\sigma_1\sigma_2^2\sigma_3 & (\rho_{13}\rho_{21}-\rho_{23})\sigma_1^2\sigma_2\sigma_3 & (1-\rho_{12}^2)\sigma_1^2\sigma_2^2 \end{pmatrix}.$$

Putting the above all partition matrix together, equation (A.4) becomes

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

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 (2.13), 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 $MSE[y_{jt}^{1:2:\ldots m}(1)] < MSE[y_{jt}^{1:2:\ldots(m-1)}(1)] <\ldots< MSE[y_{jt}^{1:2}(1)] < MSE[y_{jt}^{1}(1)]$. ■

Proof of proposition 6. Given $\sigma_1=\sigma_2=\sigma_3=1$, equation (2.18) becomes

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

Similarly,

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

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

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

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

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

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

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

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 (2.13). 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 $S^{1:2}$ to $S^{2:1}$. Since $\sigma_2 > \sigma_1$, the term one of equation (2.18) is larger than that of $S^{2:1}$. Because $\rho_{23} > \rho_{13}$, the term two of equation (2.18) is bigger than that of $S^{2:1}$. 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 expression of $S^{1:3}, S^{3:1}$ and $S^{2:3}, S^{3:2}$, 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 $S^{2:1}$ is smaller than that of expression $S^{3:1}$ since $\rho_{12} > \rho_{13}$ by assumption. Now we combine the last two terms for expression $S^{2:1}$, we have

$$\frac{2(1-\rho_{13}^2-\rho_{23}^2-\rho_{12}^2)+(\rho_{13}+\rho_{12}\rho_{23})^2}{1-\rho_{12}^2}\sigma_3^2 \tag{A.5}$$

Likewise, we combine the last two terms for expression $S^{3:1}$, we have

$$\frac{2(1-\rho_{13}^2-\rho_{23}^2-\rho_{12}^2)+(\rho_{12}+\rho_{13}\rho_{23})^2}{1-\rho_{13}^2}\sigma_2^2 \tag{A.6}$$

If we can show expression (A.5) is smaller than expression (A.6), then we can find $min\{S^{2:1}, S^{3:1}\}$. Since

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

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 (A.5) is smaller than expression (A.6). 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 $S^{2:1}$ and equation $S^{2:3}$ 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. ■

Proof of proposition 8. Suppose the true DGP is as equation (2.20) and (2.21).

We estimate the ordinary least squares model of the form

$$y_{2t} = \beta y_{1t} + u_t$$

OLS estimator ($\hat{\beta}_{ols}$) is

$$\begin{aligned}
\hat{\beta}_{ols} &= \frac{\Sigma_{t=1}^{T} y_{1t} y_{2t}}{\Sigma_{t=1}^{T} y_{1t}^2} = \frac{\Sigma_{t=1}^{T} y_{1t}(\delta y_{1t-1} + \beta y_{1t} + u_t)}{\Sigma_{t=1}^{T} y_{1t}^2} \\
&= \beta + \delta \frac{\Sigma_{t=1}^{T} y_{1t} y_{1t-1}}{\Sigma_{t=1}^{T} y_{1t}^2} + \frac{\Sigma_{t=1}^{T} y_{1t} u_t}{\Sigma_{t=1}^{T} y_{1t}^2} \\
&= \beta + \frac{\Sigma_{t=1}^{T} y_{1t} u_t}{\Sigma_{t=1}^{T} y_{1t}^2}.
\end{aligned}$$

Since y_{1t} and u_t are correlated, the second term goes to zero given the assumption of $\delta = 0$ while the third term does not go to zero in the limit and the estimator is biased, that is,

$$E(\hat{\beta}_{ols}) = \beta + E(\frac{\Sigma_t y_{1t} u_t}{\Sigma_t y_{1t}^2}).$$

This completes the proof of bias.

To see the inconsistency,

$$\begin{aligned}
plim_{t\to\infty}(\hat{\beta}_{ols}) &= plim_{t\to\infty}(\frac{\Sigma_{t=1}^{T} y_{1t} y_{2t}}{\Sigma_{t=1}^{T} y_{1t}^2}) \\
&= plim_{t\to\infty}(\frac{\sum y_{1t}(\beta y_{1t} + \alpha y_{1t-1} + u_t)}{\sum y_{1t}^2}) \\
&= \beta + plim_{t\to\infty} \frac{\sum y_{1t} u_t}{\sum y_{1t}^2} \\
&= \beta + \frac{Cov(y_{1t}, u_t)}{Var(y_{1t})} \\
&= \beta + \frac{\rho\sigma_1\sigma_2 - \beta\sigma_1^2}{\sigma_1^2} \\
&= \rho\frac{\sigma_2}{\sigma_1}.
\end{aligned}$$

Thus $\hat{\beta}_{ols}$ will underestimate β if $\beta > \rho\sigma_2/\sigma_1$ and will overestimate β if $\beta < \rho\sigma_2/\sigma_1$. ■