

Hyung, Namwon

Working Paper

Linking series generated at different frequencies and its applications

Research Notes, No. 99-1

Provided in Cooperation with:

Deutsche Bank Research, Frankfurt am Main

Suggested Citation: Hyung, Namwon (1998) : Linking series generated at different frequencies and its applications, Research Notes, No. 99-1, Deutsche Bank Research, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/40262>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Linking Series Generated at Different Frequencies and its Applications

Namwon Hyung

December 1998

Abstract

This paper systematically studies the use of mixed-frequency data sets and suggests that the use of high frequency data in forecasting economic aggregates can improve forecast accuracy. The best way of using this information is to build a single model, for example, an ARMA model with missing observations, that relates data of all frequencies. The implementation of such an approach, however, poses serious practical problems in all but the simplest cases. As a feasible and consistent alternative, we propose a two-stage procedure to obtain pseudo high frequency data and to subsequently use these artificial values as proxies for macroeconomic or financial models. This alternative method yields a sub-optimal forecast in general but avoids the computational problems of a full-blown single model. Our approach differs from classical interpolation since we only use past and current information to get the pseudo series. A proxy, which is constructed by classical interpolation, may fit very well in sample, but it is not useful for out-of-sample forecasts. As applications of linking series generated at different frequencies, we show that the use of monthly proxies of GDP improves the predictability of absolute stock returns and the unemployment rate compared to the use of industrial production as an alternative proxy.

JEL:C3; C4; C5

Keywords: Linked ARMA, Kalman Filter, Interpolation, Temporal Transformation, High Frequency Data, Virtual Reality Variable

Namwon Hyung, Department of Economics, University of California, San Diego

9500 Gilman Drive, La Jolla, CA 92093-0508

Email: nhyung@weber.ucsd.edu

Comments and suggestions from Clive, W.J. Granger are gratefully acknowledged. Furthermore, I would like to thank Christian Haefke. Any errors are my own.

Table of Contents

I. Introduction	3
II. Econometric models of linking series generated at different frequencies	4
III. Forecasts using higher frequency information	11
IV. Virtual reality variables	16
V. Conclusions and future research	28
References	30

I. Introduction

It is common to have data with different frequencies in a system of time series. GDP, for example, is reported quarterly, industrial production data are monthly, interest rates or stock prices can be collected at any interval. When observing a time series at a very short interval, that is, high frequency, it is difficult to relate the slow-moving components of the market or economy with the rapidly changing parts of the variables. Researchers usually convert the higher frequency data to the lower frequency data by temporal aggregation or by the stock-end method. A model for linking series generated at different frequencies can utilize all of the available information, which otherwise might be used inefficiently by excluding data at different frequencies. In theory, the best way to use these data is to build a single model that relates data of all frequencies. Unfortunately, though, building such a comprehensive model is very complicated (see Miller and Chin, 1996; Howrey, 1991). The implementation of such an approach poses serious practical problems in all but the simplest cases.

For small linear systems parameter estimation is relatively straightforward—at least in principle. Various methods can be used to estimate the parameters of vector ARMA models with different frequency data. The prediction error variance decomposition of the Gaussian likelihood function, combined with recursive calculations of the Kalman filter, can be used to calculate the value of the likelihood function implied by a *linked ARMA* model¹. However, it is not easy to use full information maximum likelihood to estimate linked-model parameters in a system as large as those that are commonly utilized for estimation or forecasting in macroeconomics.

It should be possible to improve the accuracy of forecasts or estimations by combining different frequency data but, nevertheless, it may not be possible to implement a *linked ARMA* model with multivariate systems of even moderate size. As a feasible and consistent alternative, we could use the *linked ARMA* model of a sub-system to get pseudo high frequency data (or virtual reality variable, hereafter VRV; Granger (1998)). In a second

step we can use these artificial values as proxies for macroeconomic or financial models. This would yield a sub-optimal estimator or forecast in general but alleviates the computational problem of a full-blown *linked ARMA* model. Consistent estimation or forecast procedures that are based on proxies for the missing data are usually computationally attractive.

The plan of the article is as follows: Section 2 discusses the temporal properties of linked ARMA process and Kalman filter estimation. Section 3 sets up the prediction using higher frequency information with an application to forecasting U.S. GDP. The VRVs of monthly output are introduced and assessed through their use in predictions of absolute stock returns and the unemployment rate in Section 4 and Section 5 concludes.

II. Econometric models of linking series generated at different frequencies

Suppose some series of k -dimensional vectors y_t^+ are not observable systematically because the sampling interval is longer than the interval of data generation, leading to temporal aggregation of flow series and systematic sampling of stock series. (Weiss, 1984) However, assume the whole y_t^+ series (including unobservable observations) are generated by a k -dimensional vector ARMA (p, q) process, which we call the *linked-ARMA model*,

$$(2.1) \quad a(L)y_t^+ = b(L)\epsilon_t,$$

where $a(L) = a_0 - a_1L^1 - \dots - a_pL^p$, $b(L) = b_0 - b_1L^1 - \dots - b_qL^q$ and the a_j and b_j are $(k \times k)$ coefficient matrices, a_0 is nonsingular and ϵ_t is vector white noise, so that $E\epsilon_t = 0$, $E\epsilon_t\epsilon_s' = \Sigma$ and $E\epsilon_t\epsilon_s' = 0$ for $t \neq s$. We will show how to construct the likelihood of a linked ARMA model, and the computation is carried out by using the Kalman filter in the next section. In the following subsection, the properties of the linked ARMA model and a transformed version of the linked ARMA model which does not have missing observations are illustrated. It is important to note the potential effect of temporal transformation on estimating, testing, and predicting the linked-ARMA model.

¹ Also called an ARMA model with missing observations. After a transformation of this model to a

II.1. Temporal transformation of a linked ARMA model

To see the effect of temporal transformation on the linked ARMA model, define a macro ARMA model (Lütkepohl, 1984). Given an integer m , we define a corresponding mk -dimensional macro ARMA (P, Q) process,

$$(2.2) \quad A(L)Y_t^+ = B(L)E_t, \text{ for } t = m\tau, \tau = 1, 2, \dots$$

$$\text{where } Y_t^+ = \begin{bmatrix} y_{mt}^+ \\ y_{m(t-1)}^+ \\ \vdots \\ y_{m(t-m+1)}^+ \end{bmatrix}, \quad E_t = \begin{bmatrix} \varepsilon_{mt} \\ \varepsilon_{m(t-1)} \\ \vdots \\ \varepsilon_{m(t-m+1)} \end{bmatrix},$$

$mk \times 1$ $mk \times 1$

$A(L) = A_0 - A_1L^1 - \dots - A_pL^p$, $B(L) = B_0 - B_1L^1 - \dots - B_qL^q$ and A_j, B_j consist of columns $jm k + 1$ through $jm k + mk$ of

$$\begin{bmatrix} a_0 & a_1 & \dots & a_p & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & a_0 & \dots & a_{p-1} & a_p & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & a_{p-2} & a_{p-1} & \dots & 0 & 0 & \dots & 0 \\ & & \dots & & & \dots & & & \dots & \\ 0 & 0 & \dots & a_0 & a_1 & \dots & a_p & 0 & \dots & 0 \end{bmatrix}, \quad \begin{bmatrix} b_0 & b_1 & \dots & b_q & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & b_0 & \dots & b_{q-1} & b_q & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & b_{q-2} & b_{q-1} & \dots & 0 & 0 & \dots & 0 \\ & & \dots & & & \dots & & & \dots & \\ 0 & 0 & \dots & b_0 & b_1 & \dots & b_q & 0 & \dots & 0 \end{bmatrix},$$

$(mk \times Pmk)$ $(mk \times Qmk)$

respectively. Here $P = \min\{n \in \mathbb{N} \mid nm \geq p\}$, and $Q = \min\{n \in \mathbb{N} \mid nm \geq q\}$.

We have to transform (2.1) to another structure because of data-availability by using (2.2). The following is a direct application of temporal and contemporaneous transformation techniques. A popular and easy but inefficient method is converting to the low frequency model by temporal transformation. For example, define a temporally transformed series y_t (aggregating y_t^+ over m periods or point sampling at the end of every m period) as

$$y_t = FY_t^+$$

where $F = l'_m \otimes I_k$, $l_m = (1, \dots, 1)'$ for temporal aggregation, $F = e'_m \otimes I_k$, $e_m = (1, 0, \dots, 0)'$ for point sampling at the end of period. This method can be easily modified to consider the

more general case where aggregated processes have different weights for the lagged values by defining the matrix F . One simple example will be presented in section 3.3. (R&R model).

Proposition 1. Let y_t^+ be as in (2.1), Y_t^+ as in (2.2) and $F \neq 0$. Then $y_t = FY_t^+$ has an ARMA (p^*, q^*) representation with $p^* \leq kp$ and $q^* \leq kp$ if $m \geq p \geq q$, $q^* \leq kp+1$ if $m \geq q \geq p$, $q^* \leq kp+Q$ otherwise.

Proof. See Lütkepohl (1984).

We could get a temporally transformed model by following this procedure. First we could rearrange (2.2) as below,

$$|A(L)|Y_t^+ = A^+(L)B(L)E_t$$

where $|a| = a^+a$ and a^+ is the adjoint of a . Premultiplying by F will produce y_t

$$|A(L)|FY_t^+ = FA^+(L)B(L)E_t, \text{ for } t = m\tau, \tau = 1, 2, \dots$$

So y_t is a VARMA (p^*, q^*) process of Proposition 1. Then the resulting model would be

$$(2.3) \quad c(l)y_t = D(l)v_t, \text{ for } t = m\tau, \tau = 1, 2, \dots$$

where $c(l) = c_0 - c_1l - \dots - c_{p^*}l^{p^*}$, $D(l) = D_0 - D_1l - \dots - D_{q^*}l^{q^*}$, $v_t \sim (0, I_k)$ and $l = L^m$.

As an example, consider a bivariate series, $y_t^+ = \begin{bmatrix} y_{1,t}^+ & y_{2,t}^+ \end{bmatrix}'$ where the series $y_{2,t}^+$ is not observed at all the points in time $t = 1, 2, \dots$. Instead, observations are available every end of quarter. For a stock variable,

$$y_{2,t} = y_{2,t}^+, t = 3\tau, \tau = 1, 2, \dots,$$

or the observations could be temporal aggregates for a flow variable,

$$y_{2,t} = \sum_{r=0}^2 y_{2,t-r}^+, t = 3\tau, \tau = 1, 2, \dots,$$

while for the value of $t \neq 3\tau$, $y_{2,t}^+$ is unobservable. Let the series $y_{1,t}^+$ be available every month, e.g., Industrial Production and the temporal aggregates of the series $y_{2,t}^+$ available every 3 month, e.g., GDP,

$$y_{2,t} = y_{2,t}^+ + y_{2,t-1}^+ + y_{2,t-2}^+, t = 3\tau, \tau = 1, 2, \dots$$

For the given a bivariate example the following temporal transformation

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ or } F = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

will produce a quarterly ARMA model from a linked ARMA model (or monthly ARMA model with missing observations). Additionally this framework permits the treatment of low and high frequency variables simultaneously in one system by appropriate choice of the matrix F . We call such a system a *mixed ARMA* model. Consider a mixed temporal transformation, i.e., transform the variable which we can not observe at the monthly frequency by the following F matrices,

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ or } F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Then the resulting model will be a *mixed ARMA* model with mixed frequency.

In the case of a temporal transformation of some variables in the system, it is possible to provide a more precise upper bound for the AR orders of y_t that are lower than those given in proposition 1. It can be obtained from the modified Wei-Weiss procedure as mentioned by Marcellino (1996).

Corollary 1. Let y_t^+ be as in (2.1). Suppose without loss of generality that the first variable needs a temporal transformation because it is not available at high frequency. Then the $m(k-1)$ vector y_t with appropriate matrix F , such as $y_t = FY_t^+ = (y_{1t}, y_{2t}, y_{2t-1}, \dots, y_{2t-m+1}, y_{3t}, \dots, y_{kt-m+1})'$, for $t = m\tau$, $\tau = 1, 2, \dots$, has an ARMA (p^*, q^*) representation with $p^* = p$, and $q^* = \min\{n \in N \mid nm \geq p(m-1) + q\}$.

Proof. Without loss of generality, consider the first equation of (2.1).

$$(2.4) \quad a_{11}(L)y_{1t}^+ + a_{12}(L)y_{2t}^+ + \dots + a_{1k}(L)y_{kt}^+ = b_1(L)\epsilon_{1t}$$

$$\text{let } a_{11}(L) = (1 - \gamma_1 L)(1 - \gamma_2 L) \dots (1 - \gamma_p L),$$

$$a_{11}^m(L) = \left(\sum_{i=0}^{m-1} \gamma_1^i L^i \right) \left(\sum_{i=0}^{m-1} \gamma_2^i L^i \right) \dots \left(\sum_{i=0}^{m-1} \gamma_p^i L^i \right), \text{ so that}$$

$$a_{11}^m(L)a_{11}(L) = (1 - \gamma_1^m L^m)(1 - \gamma_2^m L^m) \cdots (1 - \gamma_p^m L^m). \text{ Premultiplying the equation (2.4)}$$

by $a_{11}^m(L)$ gives $a_{11}^m(L)a_{11}(L)y_{1t}^+ + a_{11}^m(L)a_{12}(L)y_{2t}^+ + \cdots + a_{11}^m(L)a_{1k}(L)y_{kt}^+ = a_{11}^m(L)b_1(L)\epsilon_{1t}$.

The degrees of polynomials, denoted by $\deg(\bullet)$, are $\deg(a_{11}^m(L)a_{1j}(L)) = mp$ for $j = 1, \dots, k$,

$\deg(a_{11}^m(L)b_1(L)) = q + (m-1)p$, respectively. If we define a corresponding $m(k-1)$ -

dimensional mixed ARMA process of y_t using the similar method of defining macro ARMA

process, then we get an AR of order mp/m and an MA of order $q^* = \min\{n \in N \mid n \geq (p(m-$

$l) + q)/m\}$.

Q.E.D.

II.2. Estimation of a linked ARMA model using Kalman filtering

The first step toward estimating the model (2.1) is to cast it into a state space form so that the Kalman filter can be used to evaluate the likelihood function. The state equation describes the evolution of the unobservable state vector y_t^+ and the observation equation relates the observed variables, y_t , to the element of the state vector.

Think about the bivariate series, $y_t^+ = (y_{1,t}^+ \quad y_{2,t}^+)$ where the series $y_{2,t}^+$ is not observed at all the points in time $t = 1, 2, \dots$. Instead, observations are available every 3 months. For a stock variable

$$y_{2,t} = y_{2,t}^+, \quad t = 3\tau, \quad \tau = 1, \dots, T.$$

Or the observations could be temporal aggregates of a flow variable,

$$y_{2,t} = y_{2,t}^+ + y_{2,t-1}^+ + y_{2,t-2}^+, \quad t = 3\tau, \quad \tau = 1, \dots, T.$$

while for the value of $t \neq 3\tau$, $y_{2,t}^+$ is unobservable.

2.2.1. The simplest Case: Stock Variables

Suppose the DGP of y_t^+ follows a VAR(1) as below,

State-Equation:

$$\begin{bmatrix} y_{1,t}^+ \\ y_{2,t}^+ \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1}^+ \\ y_{2,t-1}^+ \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \quad \text{where } F = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

Observation Equation:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = H' \begin{bmatrix} y_{1,t}^+ \\ y_{2,t}^+ \end{bmatrix}, \text{ where } H' = I_2 \quad \text{if } t = m\tau, \tau = 1, \dots, T.$$

$$y_{1,t} = H' \begin{bmatrix} y_{1,t}^+ \\ y_{2,t}^+ \end{bmatrix}, \text{ where } H' = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad \text{otherwise}$$

$$\text{where } E(\varepsilon_t \varepsilon_\tau') = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}, \quad \varepsilon_t = (\varepsilon_{1,t} \quad \varepsilon_{2,t})', \text{ and } Q = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

The Kalman filter is a well-known way to compute the Gaussian likelihood function for a trial set of parameters; for a discussion, see Hamilton (1994). The filter recursively constructs minimum mean square error estimates of the unobserved state vector, given observations on y_t . The filter consists of two sets of equations, the forecast and updating equations. Let $y_{t|\tau}^+$ denote the estimate of y_t^+ based on (y_1, \dots, y_τ) , and let $P_{t|\tau}$ denotes

$$E\left[(y_{t|\tau}^+ - y_t^+)(y_{t|\tau}^+ - y_t^+)' \right]. \text{ With this notation, the forecast and the updating equations of the}$$

Kalman filter follow as below.

Forecasts and associated MSE

$$y_{t|\tau-1}^+ = F y_{t-1|\tau-1}^+, \text{ and MSE } P_{t|\tau-1} = F P_{t-1|\tau-1} F' + Q \text{ where } P_{t|\tau-1} = \begin{pmatrix} P_{t|\tau-1}^{11} & P_{t|\tau-1}^{12} \\ P_{t|\tau-1}^{21} & P_{t|\tau-1}^{22} \end{pmatrix}$$

$$\begin{bmatrix} y_{1,t|\tau-1} \\ y_{2,t|\tau-1} \end{bmatrix} = \begin{bmatrix} y_{1,t|\tau-1}^+ \\ y_{2,t|\tau-1}^+ \end{bmatrix} \text{ and MSE } G_{t|\tau-1} = P_{t|\tau-1}, \text{ if } t = m\tau, \tau = 1, \dots, T.$$

$$y_{1,t|\tau-1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t|\tau-1}^+ \\ y_{2,t|\tau-1}^+ \end{bmatrix} \text{ and MSE } G_{t|\tau-1}^{11} = P_{t|\tau-1}^{11}, \text{ otherwise}$$

Updating the Inference about y_t^+

$$y_{t|\tau}^+ = \begin{cases} y_{t|\tau-1}^+ + y_t - y_{t|\tau-1} & \text{if } t = m\tau \\ y_{t|\tau-1}^+ + \left(\begin{matrix} 1 \\ P_{t|\tau-1}^{21} (P_{t|\tau-1}^{11})^{-1} \end{matrix} \right) (y_{1,t} - y_{1,t|\tau-1}) & \text{otherwise} \end{cases}$$

$$P_{t|t} = \begin{cases} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \text{if } t = m\tau \\ \begin{pmatrix} 0 & 0 \\ 0 & P_{t|t-1}^{22} - P_{t|t-1}^{21} (P_{t|t-1}^{11})^{-1} P_{t|t-1}^{12} \end{pmatrix} & \text{otherwise} \end{cases}$$

The Kalman filter equations permit recursive calculation of the predicted state vectors and the covariance of this estimate given the assumed parameters in F and Q for given initial values for $y_{t|t}^+$ and $P_{t|t}$. If we assume that the initial state and the innovations are Gaussian, then

$$y_t | Y_{t-1} \sim N(y_{t|t-1}^+, P_{t|t-1}) \quad \text{if } t = m\tau$$

$$y_{1,t} | Y_{t-1} \sim N(y_{1,t|t-1}^+, P_{t|t-1}^{11}) \quad \text{otherwise}$$

where $Y_{t-1} \equiv (y_{t-1}', y_{t-2}', \dots)'$ and $y_{2,t}$ is not observable unless $t=m\tau$, $\tau=1,2,\dots$. That is,

$$f(y_t | Y_{t-1}) = (2\pi)^{-1} |P_{t|t-1}|^{-1/2} \exp\left\{-\frac{1}{2}(y_t - y_{t|t-1}^+) P_{t|t-1}^{-1} (y_t - y_{t|t-1}^+)\right\} \quad \text{if } t = m\tau$$

$$f(y_{1,t} | Y_{t-1}) = (2\pi)^{-1/2} |P_{t|t-1}^{11}|^{-1/2} \exp\left\{-\frac{1}{2}(y_{1,t} - y_{1,t|t-1}^+) P_{t|t-1}^{11} (y_{1,t} - y_{1,t|t-1}^+)\right\} \quad \text{otherwise}$$

The sample log likelihood is then computed as follows:

$$L = \sum_{t=m\tau, \tau=1,2,\dots} \log f(y_t | Y_{t-1}) + \sum_{t \neq m\tau, \tau=1,2,\dots} \log f(y_{1,t} | Y_{t-1})$$

The Gaussian maximum likelihood estimates of the parameters are found by maximizing L over the parameter space.

II.2.2. A Difference Stationary Process of Aggregated Flow Variables

Suppose the DGP of y_t^+ follows a difference stationary $VAR(1)$ as below,

$$\begin{bmatrix} \Delta y_{1,t}^+ \\ \Delta y_{2,t}^+ \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1}^+ \\ \Delta y_{2,t-1}^+ \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \quad \text{where} \quad \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right)$$

Since the difference of the temporally aggregated flow variable is given by,

$$\begin{aligned}
\Delta y_{2,t} &= y_{2,t} - y_{2,t-1} \\
&= (y_{2,t}^+ + y_{2,t-1}^+ + y_{2,t-2}^+) - (y_{2,t-3}^+ + y_{2,t-4}^+ + y_{2,t-5}^+) \\
&= \Delta y_{2,t}^+ + 2\Delta y_{2,t-1}^+ + 3\Delta y_{2,t-2}^+ + 2\Delta y_{2,t-3}^+ + \Delta y_{2,t-4}^+
\end{aligned}$$

its state-space form is given by:

State-Equation:

$$\begin{bmatrix} \Delta y_{1,t}^+ \\ \Delta y_{2,t}^+ \\ \Delta y_{2,t-1}^+ \\ \Delta y_{2,t-2}^+ \\ \Delta y_{2,t-3}^+ \\ \Delta y_{2,t-4}^+ \end{bmatrix} = \begin{bmatrix} a & b & 0 & 0 & 0 & 0 \\ c & d & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1}^+ \\ \Delta y_{2,t-1}^+ \\ \Delta y_{2,t-2}^+ \\ \Delta y_{2,t-3}^+ \\ \Delta y_{2,t-4}^+ \\ \Delta y_{2,t-5}^+ \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

Observation Equation:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} \Delta y_{1,t}^+ \\ \Delta y_{2,t}^+ \\ \Delta y_{2,t-1}^+ \\ \Delta y_{2,t-2}^+ \\ \Delta y_{2,t-3}^+ \\ \Delta y_{2,t-4}^+ \end{bmatrix} \quad \text{if } t = m\tau, \tau = 1, \dots, T,$$

$$\Delta y_{1,t} = \boldsymbol{\mu}_1 + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{1,t}^+ \\ \Delta y_{2,t}^+ \\ \Delta y_{2,t-1}^+ \\ \Delta y_{2,t-2}^+ \\ \Delta y_{2,t-3}^+ \\ \Delta y_{2,t-4}^+ \end{bmatrix} \quad \text{otherwise}$$

III. Forecasts using higher frequency information

When researchers have data with different frequencies in a system of time series, they usually convert the higher frequency data to the lower frequency data by temporal transformation. Such a method might be utilizing the information inefficiently by excluding data at different frequencies.

Rathjens and Robins (1993) show that quarterly forecast can be improved by using within-quarter variation of monthly data. In their paper, they did not exploit all of the within-quarter information of monthly data. Instead they used the monthly growth rate of industrial production at the end of each quarter by comparing the third month to the quarterly average. Also they use a model which can not be immediately extended to a multi-step forecast. Instead they change the specification slightly and show that within-quarter movement of monthly data is not useful for multi-step forecasts.

In this section, we will show that forecasts using all of the monthly data can predict more accurately than forecasts using part of the monthly data, and compare our results with the model of Rathjens and Robins (1993).

The following proposition states that our method will be better than Rathjens and Robins (1993).

Proposition 2: (*Forecast Efficiency of Linearly Transformed VARMA Processes*)

Let y_t be a stable, invertible, N -dimensional VARMA(p, q) process, let F be an $(M \times N)$ matrix of rank M and let $z_t = Fy_t$. Furthermore, denote the MSE matrices of the optimal h -step predictors of y_t and z_t by $\Sigma_y(h)$ and $\Sigma_z(h)$, respectively. Then

$$\Sigma_z(h) - F\Sigma_y(h)F'$$

is positive semidefinite.

Proof: See Lütkepohl(1987, Chapter 4, p.101), It can be shown that $Fy_t(h)$ is the optimal predictor based on a Hilbert space H_t generated by the $y_s, s \leq t$. Since F is an $(M \times N)$ matrix of rank M , $\mathfrak{R}(y_t) \supset \mathfrak{R}(z_t)$ where $\mathfrak{R}(\bullet)$ is a vector space. Since the $z_s, s \leq t$ generate only a subspace of H_t the predictor based on that space can not be better than $Fy_t(h)$.

Q.E.D.

These results hold for VARMA processes for which all the parameters are known. They do not necessarily carry over to estimated processes. With properly specified equations there are two sources of forecast error. The first one is due to unobserved innovations in the equation for the forecast period, and the second one is coefficient uncertainty. That is,

estimated coefficients of the equation deviate from the true coefficients in a random fashion, so the results of Proposition 2 could be not true sometimes.

III.1. Model 1: Linked ARMA

For simplicity, let $\mu_1 = \mu_2 = 0$, $Y_t^O = F_O Y_t^+$, where $t = 3\tau$, $\tau = 1, \dots, T$.

$$Y_t^O = \begin{bmatrix} \Delta y_{1,t}^m & \Delta y_{1,t-1}^m & \Delta y_{1,t-2}^m & \Delta y_{2,t}^q \end{bmatrix}$$

$$Y_t^+ = \begin{bmatrix} \Delta y_{1,t}^+ & \Delta y_{1,t-1}^+ & \Delta y_{1,t-2}^+ & \Delta y_{1,t-3}^+ & \Delta y_{1,t-4}^+ & \Delta y_{2,t}^+ & \Delta y_{2,t-1}^+ & \Delta y_{2,t-2}^+ & \Delta y_{2,t-3}^+ & \Delta y_{2,t-4}^+ \end{bmatrix}$$

$$F_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

III.2. Quarterly Model: VAR(1)

$Y_t^O = F_Q Y_t^+ = F_{Q1} Y_t^O$, where $t = 3\tau$, $\tau = 1, \dots, T$.

$$Y_t^O = \begin{bmatrix} \Delta y_{1,t}^q & \Delta y_{2,t}^q \end{bmatrix}$$

$$F_Q = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

$$F_{Q1} = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let $y_t = Y_t^O$ and $z_t = Y_t^O$, then $z_t = F_{Q1} y_t$. By Proposition 2 the Quarterly model will produce sub-optimal forecasts compared to Model I.

III.3. Rathjens and Robins' Model

Rathjens and Robins show how to improve quarterly forecasts by using within-quarter variations of monthly data. Within-quarter movements may contain valuable information for forecasting across quarter movements. They find the variable (x_τ) defined as the difference between the third month of the quarter and the average of the quarter to be useful in one-step-ahead forecasting.

$$\Delta y_{2,\tau} = \mu + a \Delta y_{2,\tau-1} + b \Delta y_{1,\tau-1} + c x_{\tau-1} + \varepsilon_{\tau}, \text{ where } \tau = 3k, k = 1, \dots, T.$$

They drop the lagged difference of GDP and industrial production from the regression because adding a new variable, x , makes these terms insignificant.

$$\Delta y_{2,\tau} = \mu + c x_{\tau-1} + \varepsilon_{\tau}, \text{ where } \tau = 3k, k = 1, \dots, T.$$

We can compare this model to our model as follows. Let

$$\begin{aligned} \tilde{\Delta} y_{1,t} &\equiv y_{1,t} - \frac{1}{3}(y_{1,t} + y_{1,t-1} + y_{1,t-2}) \\ &= \frac{2}{3} \Delta y_{1,t} + \frac{1}{3} \Delta y_{1,t-1} \end{aligned}$$

$$Y_t^R = F_R Y_t^+ = F_{R1} Y_t^O, \text{ where } t = 3\tau, \tau = 1, \dots, T.$$

$$Y_t^R = \begin{bmatrix} \tilde{\Delta} y_{1,t} & \Delta y_{2,t}^q \end{bmatrix}$$

$$F_R = \begin{bmatrix} 2/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

$$F_{R1} = \begin{bmatrix} 2/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let $y_t = Y_t^O$ and $z_t = Y_t^R$, then $z_t = F_{R1} y_t$. By Proposition 2, Rathjens and Robins' Model will produce a sub-optimal forecast compared to Model I. No other model can produce a better forecast than Model I for a given dataset in terms of MSE.

III.4. Application to GDP and IP

In this section, the empirical results of our model, a quarterly VAR model, and the Rathjens and Robins' model are compared by the one-quarter-ahead forecastability of the GDP. We convert monthly industrial production data to quarterly data by averaging within-quarter data. In the previous section we have demonstrated that monthly data can improve the accuracy of forecasts of quarterly aggregates. The usefulness of a linked ARMA model is that it can use monthly data systematically rather than in an informal way.

Table 1. Monthly MSFE for output equation in the Linked ARMA model

h	1	2	3	4	5	6	9	12
<i>MSFE</i>	0.0669	0.0872	0.1131	0.1042	0.0927	0.0957	0.0994	0.0991

(1) The full sample is 1959:10 ~ 1995:12. For the estimation, we select 1959:10 ~ 1983:03, 294 observations. All of the comparison of out-of-sample predictability uses the sample 1983:04 ~ 1995:12, 141 monthly observations

(2) h denotes a h -months ahead forecast

(3) 3, 6, 9 and 12 months ahead forecasts are identical to 1, 2, 3 and 4 quarters ahead forecasts in the table 2.

Table 1 shows different forecasting exercises were performed for each sample period to duplicate the amount of information available over the course of the quarter. It tests the forecasting capabilities of the linked ARMA model for the quarterly GDP. These GDP forecasts are only testable when a predicted month is corresponding to the end of quarter. Now consider moving through time. The information set now available is previous information and the current month's information. We are in the second month of the quarter, when we have one more observation for monthly series, so it is 2 or 5-month-ahead forecasts. Finally we do 1 or 4 month-ahead forecasts at the end of the second month of each quarter. For the short horizon (one or two month ahead forecasts), the use of within-quarter information clearly improves predictability as seen in table 1.

Table 2. In-sample Comparison for output equation

	<i>Linked ARMA</i>	<i>Quarterly VAR(1)</i>	<i>Rathjens & Robin</i>
<i>SSR</i>	6.0018	14.9557	11.6136
<i>DW</i>	2.0946	1.8045	1.8475
<i>AIC</i>	-2.5888	-1.8078	-2.0607
<i>BIC</i>	-2.3258	-1.7281	-1.9811
<i>Prob(F-stat)</i>	-	0.0007	0.0000
<i>Loglik</i>	-2.2031	-46.9604	-34.6940

The full sample is 1959:10 ~ 1995:12. For the estimation, we select 1959:10 ~ 1983:03, 294 monthly observations of Industrial production and 98 quarterly observations of GDP.

Table 3. Out-of-sample Comparison (MSFE) for output equation

	<i>Linked ARMA</i>	<i>Quarterly VAR(1)</i>	<i>Rathjens & Robin</i>	<i>Link/VAR</i>	<i>Link/R&R</i>
$h=1$	0.1131	0.0959	0.1158	1.18	-0.83
$h=2$	0.0957	0.0889	0.0908	1.45	1.02
$h=3$	0.0994	0.0952	0.0962	1.08	0.83
$h=4$	0.0991	0.0966	0.0994	1.26	-0.20

- (1) The full sample is 1959:10 ~ 1995:12. For the estimation, we select 1959:10 ~ 1983:03, 294 observations. All of the comparison of out-of-sample predictability uses the sample 1983:04 ~ 1995:12, 47 quarterly observations
- (2) h denotes a h -quarters ahead forecast
- (3) For comparison, select quarterly aggregated observations and do monthly forecasts and aggregated these predicted values to get predicted quarterly aggregated values in the *Linked ARMA* model
- (4) Columns 4 and 5 show test statistics of predictability between two different models, which follow standard normal distributions. See Granger and Newbold (1976) and Meese and Rogoff (1987).

The next two tables show the comparison between three methods of combining two different series. Table 2 shows the in-sample comparison and table 3 presents the mean squared forecast error of out-of-samples for these 3 methods. A linked ARMA method improves the in-sample explanation but no significant improvement of out-of-sample forecastability is evident. For the short forecast horizon, this method shows a marginal but not significant improvement over Rathjens and Robins' method. None of these methods can beat the other method in terms of out-of-sample predictability.

IV. Virtual reality variables

To get a longer time series, one can observe the same time span but at a higher frequency rate, going from quarterly to monthly². Observing data more frequently clearly provides more data but it does not necessarily provide more information. However, in applied work several estimators have been proposed which are based on approximations substituting for the missing data (Bernanke, Gertler and Watson (1997), Bernanke and Mihov (1995). Frequently the model is too large to be jointly estimated or it is only specified in part. It may thus be necessary to use a consistent estimation method instead of a fully efficient estimator to reduce the problem of complexity. Consistent procedures that are based on proxy variables

² See Granger (1998) for an extensive discussion of this topic

for the missing figures are usually computationally attractive. In this section we will show how such consistent estimates using proxy variables can be obtained.

Consider several methods of constructing pseudo high frequency data. The first approach consists of assuming a univariate model, which could be interpreted as the marginal processes of some linear simultaneous system (see Zeller and Palm 1974). It is proposed to avoid a considerable loss of information which is induced by aggregation to get consistent frequencies of the data. For example, Chan (1993) did not consider the case which disaggregates annual series to quarterly series without any additional information from other related series.

Second, related variables that are observed more frequently can be included in the conditioning set of the expectation of the missing variables. The regression model analyzed by Chow and Lin (1971) or Palm and Nijman (1984) will be an appropriate tool for estimating the missing observations. It is likely that even though the monthly conditional expectation of a variable which is observed quarterly, given its information set, is not identified, those of its conditional expectation given some indicator variable may be identified. An indicator variable will be required for estimating the missing values. In their method, they are using interpolation to get missing data rather than prediction. The third method, which we will focus on in this section, is using the same technique but using an information set that is different from the one used in the second method.

Consider the following example to illustrate the third method. Let one variable be measured quarterly in a model of variables which are otherwise all observed monthly. Suppose that a VAR type of model is being considered relating the vector of monthly series and a vector of a quarterly series. We could envisage a model that exists monthly, estimating a value for the process each month and then, with each month as a base, forecasting the missing monthly value of the quarterly series. The estimated "monthly" values of quarterly series can be thought of as being a virtual reality variable (VRV) of the "monthly" values of the quarterly series. The idea can obviously be extended to higher frequencies. Few of the virtual GDP figures are actually observed, they are estimated from some model using

previous values and other data of similar frequency but can be compared to the actual value when it is observed each quarter. The model will need to contain an "error-correction" term in case the predicted GDP figures deviate from the observed values to bring the sequence back onto course.

In this section we focus on the estimation and forecasting properties of the VRV method and present some empirical evidence for the VRV approach forecasting absolute stock returns and the unemployment rate.

IV.1. The Model, Assumptions, and Notation

In matrix notation, the model considered is

$$(4.1) \quad y = x\beta + \varepsilon$$

$$(4.2) \quad x = z\pi + \nu$$

where (4.1) is the structural equation of interest, y and x are $T \times 1$ vectors of T observations, (4.2) is the reduced form equation where the observations on the incompletely observed variables, x , are regressed on auxiliary variables, z , $T \times 1$ vector. The regression disturbances ε_t and ν_t are i.i.d. with mean zero and variances σ_ε^2 and σ_ν^2 respectively, have finite fourth moments, are independent of the corresponding regressors and satisfy $E[\varepsilon_t \nu_\tau] = 0$ for all t, τ .

Further assume,

$$(4.3) \quad z'z/T \xrightarrow{p} Q, \text{ where } Q \text{ is finite and non-singular.}$$

$$(4.4) \quad E[z_t \varepsilon_\tau] = 0 \text{ for all } t, \tau.$$

To show how consistent proxy variable estimators can be obtained, consider equation (4.1). Define $X_t = (y_t, x_t)'$. Assume that the conditional expectation of X_t given some information set z_t exists, and define,

$$(4.5) \quad \tilde{X}_t = E[X_t | z_t, \pi] \text{ and } \hat{X}_t = E[X_t | z_t, \hat{\pi}]$$

where π and $\hat{\pi}$ are vectors of parameters in the conditional expectations and their estimate, respectively.

When X_t is not observed for some or all t , \hat{X}_t can be used as proxy for X_t , provided a consistent estimate $\hat{\pi}$ is available. The proxy equals the observed value, whenever the latter is available. Substitution of \hat{X}_t into (4.1) yields

$$(4.6) \quad y_t = \hat{x}_t \beta + u_t$$

where \hat{x}_t is the appropriate element of \hat{X}_t , and $u_t = \varepsilon_t + (x_t - \hat{x}_t)\hat{\pi}$ since y_t is observable for all the time periods.

Proposition 3: Suppose that (4.1) and (4.2), and assumptions (4.3) and (4.4) hold. Then ordinary least squares (OLS) applied to (4.6) will be consistent for β .

Proof: It suffices to show that

$$\left(\sum_{t=1}^T \hat{x}_t^2 \right)^{-1} \sum_{t=1}^T \hat{x}_t u_t \xrightarrow{p} 0.$$

With a little algebra,

$$\begin{aligned} \sum_{t=1}^T \hat{x}_t u_t &= \sum_{t=1}^T z_t \hat{\pi} \{ \varepsilon_t - \hat{\pi}(z_t \pi + v_t - z_t \pi) \} \\ &= \hat{\pi} \sum_{t=1}^T z_t \varepsilon_t - \hat{\pi}^2 \sum_{t=1}^T z_t v_t + \hat{\pi}(\hat{\pi} - \pi) \sum_{t=1}^T z_t^2 \end{aligned}$$

From (4.5), (4.4) and if a consistent estimate $\hat{\pi}$ is available, then

$$\sum_{t=1}^T \hat{x}_t u_t \xrightarrow{p} 0. \quad \text{Q.E.D.}$$

Variables that are independent of ε_t are also asymptotically independent of u_t , provided they are included in the conditioning set of (4.6). If we can calculate a consistent proxy \hat{x}_t from (4.2) under the usual conditions of the two-stage least squares structure of IV estimation, then OLS applied to (4.6) will yield a consistent estimate. In this case, a consistent proxy \hat{x}_t is an appropriate instrumental variable to estimate β consistently from (4.2).

Next, we will show that the VRV procedure can forecast better than monthly or quarterly VAR procedures by using proposition 2 in section 3. Under the previous assumptions, we can compare the forecastability of various models. Let

$y_t^+ = [y_{1t} \ y_{1t-1} \ y_{1t-2} \ y_{2t}^+ \ y_{2t-1}^+ \ y_{2t-2}^+ \ y_{3t} \ y_{3t-1} \ y_{3t-2}]$, where $+$ denotes unobservable variables and $t = 3\tau$, $\tau = 1, \dots, T$. For example, let y_1 be stock returns, y_2 be GDP and y_3 be the Industrial production index which is an auxiliary variable in this system.

(1) Linked ARMA model:
$$F = \begin{bmatrix} I_3 & 0 \\ & l'_3 \\ 0 & I_3 \end{bmatrix} \text{ where } l_3 = [1 \ 1 \ 1]'$$

(7×9)

(2) Quarterly model 1: Use quarterly aggregated data of three variables

$$F = I_3 \otimes l'_3$$

(3×9)

(3) Quarterly model 2: Use quarterly aggregated data not including auxiliary variables

$$F = \begin{bmatrix} l'_3 & 0 & 0 \\ 0 & l'_3 & 0 \end{bmatrix}$$

(2×9)

(4) Monthly model: Use monthly proxy instead of quarterly aggregated data

$$F = \begin{bmatrix} I_3 & 0 & 0 \\ 0 & 0 & I_3 \end{bmatrix}$$

(6×9)

(5) VRV model: Generate monthly proxy (VRV) and do VARMA regression using VRV, where VRV is calculated by a linear projection using the Kalman Filter method.

Proposition 4: (*Forecast Efficiency of VRV Method*) Let y_t^+ be a stable, invertible VARMA process with missing observations. Then a linked ARMA and VRV model can produce better forecasts than the quarterly or monthly models introduced above.

Proof: Show the equivalent predictability of linked ARMA and VRV model under certain conditions by proving $\mathfrak{R}(Y_t) = \mathfrak{R}(\hat{Y}_t)$, where $Y_t = (y_{1t} \ y_{1t-1}, \dots, y_{2t} \ y_{2t-3}, \dots, y_{3t} \ y_{3t-1}, \dots)$,

$\hat{Y}_t = (y_{1t}, y_{1t-1}, \dots, \hat{y}_{2t}, \hat{y}_{2t-1}, \dots, y_{3t}, y_{3t-1}, \dots)$. \hat{y}_{2t} is a linear projection of y_{2t}^+ on $\mathfrak{R}(I_t)$

where $I_t = (y_{2t}, y_{2t-3}, \dots, y_{3t}, y_{3t-1}, \dots)$, \hat{y}_{2t+1} is a linear projection of y_{2t+1}^+ on $\mathfrak{R}(J_{t+1})$ where

$J_t = (y_{2t}, y_{2t-3}, \dots, y_{3t+1}, y_{3t}, \dots)$ and \hat{y}_{2t+2} is a linear projection of y_{2t+2}^+ on $\mathfrak{R}(H_t)$ where $H_t = (y_{2t}, y_{2t-3}, \dots, y_{3t+2}, y_{3t+1}, \dots)$. Since \hat{y}_{2t} is a linear combination of its past available information, \hat{Y}_t is only the linear combination of Y_t . By the projection theorem (Theorem 7.6.5 of Anderson (1971)) for Hilbert spaces, the optimal h -step predictors of y_t based on a Hilbert space generated by Y_t is unique and equivalent to the optimal h -step predictors of y_t based on a Hilbert space generated by \hat{Y}_t .

It is easy to show that a linked ARMA model produces better forecasts than the quarterly or monthly model. Let $y_t = (y_{1t}, y_{1t-1}, y_{1t-2}, y_{2t}, y_{3t}, y_{3t-1}, y_{3t-2})'$ where $y_{2t} = y_{2t}^+ + y_{2t-1}^+ + y_{2t-2}^+$, then F can be defined as the following: for the quarterly model 1,

$$F = \begin{bmatrix} I'_3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I'_3 \end{bmatrix}, \text{rank } F = 3. \text{ For the quarterly model 2, } F = \begin{bmatrix} I'_3 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{rank } F = 2. \text{ For}$$

the monthly model, $F = \begin{bmatrix} I_3 & 0 & 0 \\ 0 & 0 & I_3 \end{bmatrix}$, rank $F = 6$. By Proposition 2, all of these

transformations produce less efficient forecast than a linked ARMA model. Q.E.D.

Using proposition 4, we could conjecture that a VRV method using the FMGDP of the following sub-section may produce a better forecast than any other model. In the empirical study of section 4.3 we compare the forecastability of the method using various VRVs and the method using industrial production as a monthly proxy of output. Also, in practice, we use a two-stage estimation method which is introduced in this section, i.e., first, construct the VRV using an auxiliary variable, then estimate the structural equation by using the VRV. This method will produce the same result under the assumption of (4.4).

IV.2. Virtual Reality Variables: Examples

The properties of estimated missing observations depend on the information set of the conditional expectation, whether it includes its past, current or future information. The

following methods have different information sets using the Kalman Filters of section 2, which are applied to calculate monthly proxies of GDP using the Industrial production index.

a. Interpolation: Interpolation is a technique often used to increase data frequency or to estimate missing observations. It involves generating new data points connecting two available and consecutive data points with additional information available from other related series. The problem with this approach is that one is using the future information to get pseudo-high-frequency data. Therefore interpolated data cannot be used for true out of sample forecasting.

$$\text{e.g., Monthly Change of GDP (SMGDP)} = \frac{\mu_2}{9} + \Delta y_{2,t|T}^+$$

b. Filtering: This is an inference about the value of current pseudo-high-frequency data using the current and past information with additional information available from other related series.

$$\text{e.g., Monthly Change of GDP (FMGDP)} = \frac{\mu_2}{9} + \Delta y_{2,t|t}^+$$

c. 1-step Prediction: Using the past information only, 1-step forecast of pseudo-high-frequency data can be calculated.

$$\text{e.g., Monthly Change of GDP (PMGDP)} = \frac{\mu_2}{9} + \Delta y_{2,t|t-1}^+$$

IV.3. Empirical examples

To illustrate the usage of VRVs, we forecast the absolute stock returns and unemployment using output. We start the empirical analysis by constructing a VRV using the system which is introduced in section 2.2. A brief analysis of estimates of monthly figures of U.S. output follows in table 4 and we plot the series of industrial production and VRVs in Figure 1. The plot of PMGDP appears to be smoother than the plots of other proxies. FMGDP shows the highest correlation with industrial production in table 4. In the following subsection we use various proxies of output to explain the model.

Table 4. Correlation Matrix between VRV and Industrial Production

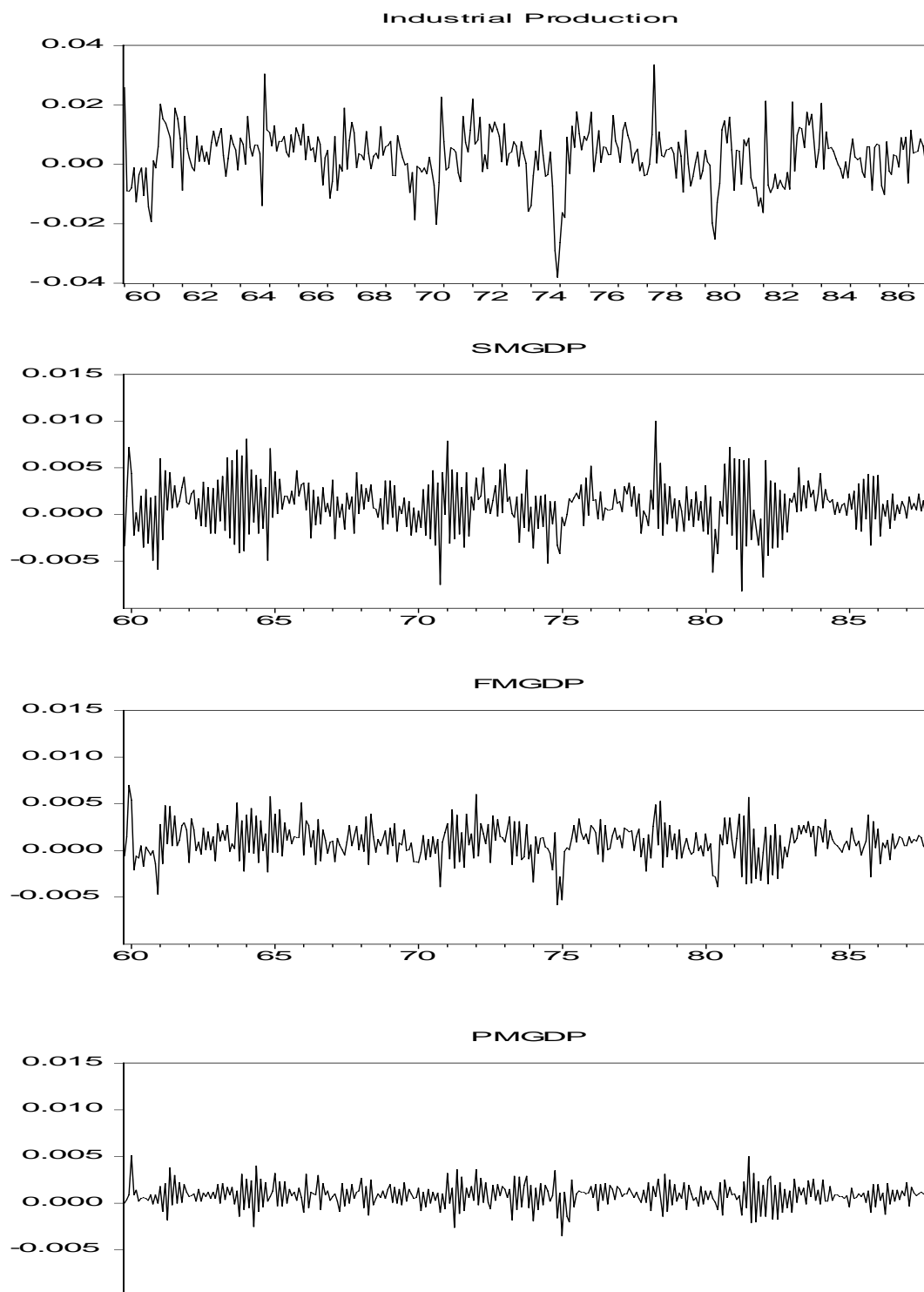
	<i>IP</i>	<i>SMGDP</i>	<i>FMGDP</i>	<i>PMGDP</i>
<i>IP</i>	1.0000	0.5065	0.6711	0.0964
<i>SMGDP</i>	0.5065	1.0000	0.7281	0.4284
<i>FMGDP</i>	0.6711	0.7281	1.0000	0.6131
<i>PMGDP</i>	0.0964	0.4284	0.6131	1.0000

IP: Difference of log industrial production (1959:10 ~ 1997:03)

SMGDP: Monthly change of log GDP estimated by interpolation

FMGDP: Monthly change of log GDP estimated by filtering

PMGDP: Monthly change of log GDP estimated by 1-step ahead forecast
GDP (1959:IV ~ 1997:I)

Figure 1. Monthly Proxies of the Changes of GDP

(a) Output and Absolute Stock return Dynamics in the U.S.

It is now well established that the stock market returns themselves contain little serial correlation which is in agreement with the efficient market hypothesis. But this empirical fact does not necessarily imply that returns are independently and identically distributed as many theoretical financial models assume. It is possible that the stock market series is serially uncorrelated but shall exhibit dependence. For example, studies using ARCH models with daily or lower frequency returns have provided strong evidence of a high degree of intertemporal volatility dependence. Granger and Ding (1994) show that absolute stock market returns³ have a long memory property.

Also we have some evidence of macroeconomic news on market volatility (Bollerslev, Cho and Kroner, 1992). A rapidly growing body of research documents forecastable components in security returns. Asset prices are commonly believed to react sensitively to economic news. Since aggregate output is serially correlated and hence predictable, the theory suggests that stock returns can be predicted based on forecasts of output. In this section we will investigate the predictable components of absolute returns using output.⁴

Table 5 shows a comparison of the fit in sample using different monthly proxies of GDP. As expected, SMGDP and FMGDP can explain the absolute stock returns better than industrial production since these two proxies are constructed by using monthly industrial production and quarterly GDP together. Since SMGDP is also using future information in its derivation, it is not surprising that it has the best in-sample explanatory power. PMGDP is the worst one, even worse than IP, since this proxy is constructed from one-month past IP information and past GDP.

³ Granger and Ding (1993) suggest absolute stock returns instead of the variance be used as a measure of risk.

⁴ If we use returns instead of absolute returns, all monthly proxies of output have similar forecastability.

Table 5. In-sample Comparison for absolute stock return equation in the bivariate system

	<i>IP</i> (3)	<i>SMGDP</i> (3)	<i>FMGDP</i> (3)	<i>PMGDP</i> (4)
R^2	0.0451	0.0681	0.0514	0.0409
\bar{R}^2	0.0294	0.0528*	0.0358	0.0197
<i>SSR</i>	0.2940	0.2870*	0.2921	0.2937
<i>DW</i>	1.9972	1.9995	1.9944	1.9960
<i>AIC</i>	-7.1053	-7.1297	-7.1119	-7.0929
<i>BIC</i>	-7.0315	-7.0559	-7.0381	-6.9979
<i>F-stat</i>	0.0095	0.0002	0.0036	0.0544
<i>Loglik</i>	800.74	805.27*	801.96	798.30

The full sample is 1959:10 ~ 1997:03. Select 1959:10 ~ 1990:12 (375) as in-sample.

The numbers in psrentheses denote lags of the bivariate VAR chosen by BIC

* denotes the most preferred model

Table 6 contains some diagnostics associated with these models. Residuals are tested against a fourth-order ARCH using the LM test and checked with the Jarque-Bera normality test. The skewness and excess kurtosis of the residuals are also reported. As pointed out by Granger and Ding (1993), the marginal distribution of absolute returns is exponential, skewness is 2 and kurtosis is 9 regardless of the value of the parameter. The residuals of the models are not normal but are close to an exponential distribution.

Table 6. Test statistics and *p*-values of ARCH and Jarque-Bera normality tests, and skewness, excess kurtosis measure of residuals from the estimated models using various GDP proxies

	<i>IP</i>	<i>SMGDP</i>	<i>FMGDP</i>	<i>PMGDP</i>
<i>ARCH-LM(4)</i>	2.60	2.60	2.78	2.52
<i>p-value</i>	0.62	0.63	0.60	0.64
<i>Skewness</i>	2.02	1.96	2.00	2.02
<i>Kurtosis</i>	11.60	11.59	11.51	11.59
<i>Jarque-Berra</i>	1374.13	1357.54	1344.03	1359.22
<i>p-vau</i> e	0.00	0.00	0.00	0.00

Another way of evaluating the estimated model using different proxies is post-sample forecasting, although the insight to be gained depends on what happens in the time series during the prediction period. The forecasts were made without re-estimating the model during

the prediction period. In all cases the MSFE of 1, 3, 6 and 12-month ahead forecasts were calculated. Because of the construction method of SMGDP, the forecast using this series is not available even though it has the best in-sample fit. From table 7, the use of VRV may improve the predictability of absolute stock returns. PMGDP provides a clear improvement and FMGDP has marginal improvement against monthly industrial production alone. Since IP alone cannot reflect all of the economy and GDP is a wider index, combined information can improve predictability as well as the in-sample fit.

Table 7. Out-of-sample Comparison for absolute stock return equation in the VAR

$h=1$	<i>IP</i>	<i>SMGDP</i>	<i>FMGDP</i>	<i>PMGDP</i>
<i>Bias</i>	-0.006312	-	-0.006816	-0.005155
<i>MAD</i>	0.015577	-	0.015788	0.015061
<i>MSFE</i>	0.000389	-	0.000393	0.000372*
$h=3$				
<i>Bias</i>	-0.006513	-	-0.006651	-0.005600
<i>MAD</i>	0.015677	-	0.015708	0.015250
<i>MSFE</i>	0.000393	-	0.000390	0.000373*
$h=6$				
<i>Bias</i>	-0.007476	-	-0.007460	-0.006949
<i>MAD</i>	0.015675	-	0.015670	0.015500
<i>MSFE</i>	0.000385	-	0.000384	0.000375*
$h=12$				
<i>Bias</i>	-0.007845	-	-0.007832	-0.007494
<i>MAD</i>	0.016033	-	0.016032	0.015891
<i>MSFE</i>	0.000395	-	0.000394	0.000390*

(1) The full sample is 1959:10 ~ 1997:03. For the estimation, we select 59:10 ~ 90:12, 375 observations. All of the comparison of out-of-sample predictability uses the sample 91:01 ~ 97:03, 75 observations

(2) h denotes a h -month ahead forecast

(3) * denotes the best model in terms of MSFE

(4) Forecast comparison using Granger-Newbold, Meese-Rogoff methods

IP vs Filtered: -0.85(1-step), 0.66(3-step), 0.85(6-step) and 0.52(12-step)

IP vs Predicted: 1.86(1-step), 2.15(3-step), 1.97(6-step) and 1.06(12-step)

Filtered vs Predicted: 1.68(1-step), 1.76(3-step), 1.88(6-step) and 1.07(12-step)

(b) Output and Unemployment Dynamics in the US

A bivariate VAR model to describe output-unemployment dynamics is used in this section. We examine the advantage of VRV over industrial production in this system. Since GDP figures are not available on a monthly basis, researchers usually use an index of industrial production as their measure of real economic activity. A natural question that emerges is whether their analyses are sensitive to the use of different output data. However,

GDP is simply a better indicator of output than industrial production since the industrial production index does not include the activity of the service sector⁵ (see Christiano and Eichenbaum, 1987, p79). Like GDP, employment measures reflect activity in almost all sectors of the economy⁶ Therefore monthly industrial production is not a good measure of output for output-unemployment dynamics.

Table 8 shows the in-sample comparison of the unemployment equation using different proxies of monthly output. SMGDP and FMGDP are better indexes than IP. PMGDP is not better than IP because of the information set which was used in the construction of the proxy. In the comparison of out-of-sample forecastability, table 10, the uses of VRV have clear gains for short horizons but not for long horizons.

Table 8. In-sample Comparison for unemployment equation in the bivariate system

	<i>IP</i> (3*)	<i>SMGDP</i> (5)	<i>FMGDP</i> (5)	<i>PMGDP</i> (5)
R^2	0.9892	0.9899	0.9895	0.9877
\bar{R}^2	0.9890	0.9896	0.9892	0.9874
<i>SSR</i>	9.8574	9.1641*	9.5270	11.1528
<i>DW</i>	2.0156	1.9733	1.9411	2.0609
<i>AIC</i>	-3.4099	-3.4503	-3.4115	-3.2702
<i>BIC</i>	-3.3259	-3.3177	-3.2789	-3.1620
<i>F-stat</i>	0.0000	0.0000	0.0000	0.0000
<i>Loglik</i>	96.2368	105.9267*	99.9070	76.2296

The full sample is 1959:10 ~ 1997:03. Select 1959:10 ~ 1985:12 as in-sample period

The numbers in parentheses denote lags of the bivariate VAR chosen by BIC

* denotes the most preferred model

Table 9. Test statistics and p -values of ARCH and Jarque-Bera normality tests, and skewness, excess kurtosis measure of residuals from the estimated models using various GDP proxies

	<i>IP</i>	<i>SMGDP</i>	<i>FMGDP</i>	<i>PMGDP</i>
<i>ARCH-LM(4)</i>	7.20	6.23	10.82	9.29
<i>p-value</i>	0.126	0.182	0.029	0.054
<i>Skewness</i>	0.18	0.09	0.36	0.41
<i>Kurtosis</i>	3.64	3.80	3.74	4.14
<i>Jarque-Bera</i>	6.89	8.44	13.18	24.85
<i>p-value</i>	0.032	0.015	0.001	0.000

⁵ The only industry included is total manufacturing, mining, and utilities.

⁶ ***Business Covered:*** Industrial - mining, manufacturing; electricity, gas and water; construction; wholesale and retail trade; transport, storage, and communication; finance, insurance, real estate, and business service; community, social, and personal services; local, state, and federal government.

Not included: Agriculture, hunting, forestry, fishing, private household, and the military.

Table 10. Out-of-sample Comparison for unemployment equation in the VAR

$h=1$	<i>IP</i>	<i>SMGDP</i>	<i>FMGDP</i>	<i>PMGDP</i>
<i>Bias</i>	-0.027265	-	-0.034210	-0.009511
<i>MAD</i>	0.119640	-	0.115031	0.119512
<i>MSFE</i>	0.022856	-	0.020871*	0.022441
$h=3$				
<i>Bias</i>	-0.065524	-	-0.077901	-0.030582
<i>MAD</i>	0.195516	-	0.184105	0.183185
<i>MSFE</i>	0.056974	-	0.050701*	0.052422
$h=6$				
<i>Bias</i>	-0.122193	-	-0.139128	-0.069888
<i>MAD</i>	0.323407	-	0.314386	0.292311
<i>MSFE</i>	0.143687	-	0.134976	0.126823*
$h=12$				
<i>Bias</i>	-0.224345	-	-0.229381	-0.158275
<i>MAD</i>	0.537131	-	0.494063	0.494589
<i>MSFE</i>	0.390404	-	0.341613*	0.344671

(1) The full sample is 1959:10 ~ 1997:03. For the estimation, we select 1959:10 ~ 1985:12, 315 observations. All of the comparison of out-of-sample predictability uses the sample 1986:01 ~ 1997:03, 135 observations

(2) h denotes a h -month ahead forecast

(3) * denotes the best model in terms of MSFE

(4) Forecast comparison using Granger-Newbold, Meese-Rogoff methods

IP vs Filtered: 2.14(1-step), 1.82(3-step), 0.64(6-step) and 0.98(12-step)

IP vs Predicted: 0.35(1-step), 1.04(3-step), 0.54(6-step) and 1.23(12-step)

Filtered vs Predicted: -1.11(1-step), -0.33(3-step), 0.54(6-step) and -0.07(12-step)

As a summary, both models confirm the usefulness of VRV against IP alone. When we use stock returns instead of absolute stock returns, the evidence is much weaker because of the properties of stock returns themselves. By the same reasoning as for the absolute stock returns, VRV is a better proxy of monthly output in the output-unemployment dynamics. Since unemployment measures reflect activity in almost all sectors of the economy, VRV is a better measure than IP. The usefulness of VRV will depend on what kind of data is used in the analysis and also on whether we can construct a better VRV for the specific purpose of research by adding additional information.

V. Conclusions and future research

We consider the use of mixed frequency data sets in a systematic way and find that incorporating high frequency data into forecasting models of economic aggregates potentially

improves forecast accuracy. The most efficient method of dealing with mixed frequency data is building and estimating one single, highly complex model. However, such an approach usually suffers from numerous numerical difficulties, such as poor convergence and high dependence on starting values. In this paper we show that a two-stage procedure using proxies that are based only on current and past information has far better numerical properties at the expense of only a minor loss in efficiency. Contrary to standard interpolation these proxies are constructed without the use of future information and can therefore be incorporated into a forecasting model.

These results hold for VARMA processes for which all the parameters are known, but do not necessarily carry over to estimated processes. In the application, we conjecture that this also holds for the case of estimated processes. We need to investigate the case of estimated processes further. Also one could extend this procedure to higher frequency data, for example, weekly or daily GDP series which might be more useful for the real economic world.

References

- Bernanke, B.S., Mark G. and M. Watson, 1997, Systematic Monetary Policy and the Effects of Oil Price Shocks, *Brookings Papers on Economic Activity*, 1, 91-157.
- Bernanke, B.S., and I. Mihov, 1995, Measuring Monetary Policy, Working Paper 5145. Cambridge, Mass.: National Bureau of Economic Research (June)
- Chan, W.-S., 1993, Disaggregation of annual time-series data to quarterly figures: A comparative study, *Journal of Forecasting*, 12, 677-688.
- Chow, G.C., and A.-L. Lin, 1971, Best linear unbiased interpolation, distribution, and extrapolation of time series by related series, *Review of Economics and Statistics*, 53, 372-375.
- Christiano, L.J. and M. Eichenbaum, 1987, Temporal aggregation and structural inference in macroeconomics, *Carnegie-Rochester Conference Series on Public Policy*, 26, 63-130.
- Granger, C.W.J., 1998, Extracting information from mega-panels and high frequency data, UCSD Working Paper.
- Granger, C.W.J., and Z. Ding, 1993, Varieties of long memory models, UCSD Working Paper.
- Granger, C.W.J., and Z. Ding, 1993, Some properties of absolute return an alternative measure of risk, UCSD Working Paper.
- Granger, C.W.J., and Z. Ding, 1994, Stylized facts on the temporal and distributional properties of daily from speculative markets, UCSD Working Paper.
- Hamilton, J.D., 1994, *Time Series Analysis*, Princeton: Princeton University Press.
- Howrey, E.P., 1991, New methods for using monthly data to improve forecast accuracy, *Comparative Performance of US Economic Models*, ed. by L.R. Klein, 227-247.
- Lütkepohl, H., 1984, Linear aggregation of vector autoregressive moving average processes, *Economics Letters*, 14, 345-350.
- Lütkepohl, H., 1987, *Forecasting Aggregated Vector ARMA Process*, New York: Springer-Verlag.

- Marcellino, M., 1996, Some temporal aggregation issues in empirical analysis, UCSD working paper.
- Miller, P.J., and D.M. Chin, 1996, Using monthly data to improve quarterly model forecasts, *Quarterly Review*, Federal Reserve Bank of Minneapolis, 16-33.
- Nijman, T.E. and F.C. Palm, 1986, The construction and use of approximations for missing quarterly observations: A model-based approach, *Journal of Business & Economic Statistics*, 4, 47-58.
- Nijman, T.E. and F.C. Palm, 1988, Consistent estimation of regression models with incompletely observed exogenous variables, *Annales d'Economie et de Statistique*, 12, 151-175.
- Palm, F.C., and T.E.Nijman, 1982, Linear regression using both temporally aggregated and temporally disaggregated data, *Journal of Econometrics*, 19, 333-343.
- Rathjens, P. and R.P. Robins, 1993, Forecasting quarterly data using monthly information, *Journal of Forecasting*, 12, 321-330.
- Weiss, Andrew A., 1984, Systematic sampling and temporal aggregation in time series models, *Journal of Econometrics*, 26, 271-281.

Recent Publications:

A. Gottschling, T. Trimbur,

“A new Approach to the Evaluation and Selection of Leading Indicators”,

RN-98-1, March 1998

A.Gottschling, C. Kreuter, P.Cornelius

“The Reaction of Exchange Rates and Interest Rates to News Releases”,

RN-98-2, June 1998

Dong Heon Kim

“Another Look at Yield Spreads: Monetary Policy and the Term Structure of Interest Rates“,

RN-98-3, October 1998

Jeffrey Sachs

“Creditor Panics: Causes and Remedies“,

RN-98-4, November 1998

© 1999. Publisher: Deutsche Bank AG, DB Research, D-60272 Frankfurt am Main, Federal Republic of Germany, editor and publisher, all rights reserved. When quoting please cite „Deutsche Bank Research“.

The information contained in this publication is derived from carefully selected public sources we believe are reasonable. We do not guarantee its accuracy or completeness, and nothing in this report shall be construed to be a representation of such a guarantee. Any opinions expressed reflect the current judgement of the author, and do not necessarily reflect the opinion of Deutsche Bank AG or any of its subsidiaries and affiliates. The opinions presented are subject to change without notice. Neither Deutsche Bank AG nor its subsidiaries/affiliates accept any responsibility for liabilities arising from use of this document or its contents. Deutsche Bank Securities Inc. has accepted responsibility for the distribution of this report in the United States under applicable requirements. Deutsche Bank AG London and Morgan Grenfell & Co., Limited, both being regulated by the Securities and Futures Authority, have respectively, as designated, accepted responsibility for the distribution of this report in the United Kingdom under applicable requirements.