Discussion Paper Series

No.169

Forecasting in large cointegrated processes

Hiroaki Chigira
Taku Yamamoto

June 2006

Hitotsubashi University Research Unit
for Statistical Analysis in Social Sciences
A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

FORECASTING IN LARGE COINTEGRATED PROCESSES *

HIROAKI CHIGIRA† TAKU YAMAMOTO

Department of Economics
Hitotsubashi University

June 2006

Abstract

It is widely recognized that taking cointegration relationships into consideration is useful in forecasting cointegrated processes. However, there are a few practical problems when forecasting large cointegrated processes using the well-known vector error correction model. First, it is hard to identify the cointegration rank in large models. Second, since the number of parameters to be estimated tends to be large relative to the sample size in large models, estimators will have large standard errors, and so will forecasts. The purpose of the present paper is to propose a new procedure for forecasting large cointegrated processes, which is free from the above problems. In our Monte Carlo experiment, we find that our forecast gains accuracy when we work with a larger model as long as the ratio of the cointegration rank to the number of variables in the process is high.

Key words: Forecasting; Cointegration; Large Models.

JEL classifications: C12, C32

---

*A previous version of this paper was presented at seminars at the European University Institute and the University of Melbourne.

†Corresponding author: Hiroaki Chigira, Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, JAPAN. E-mail: hchigira@mercury.ne.jp. The author gratefully acknowledges financial support from a JSPS Research Fellowship for Young Scientists.

1 INTRODUCTION

One of the main objectives of applying cointegration analysis in forecasting is whether imposing cointegration improves the accuracy of forecasts. Using Monte Carlo experiments, Engle and Yoo (1987) showed that their long-run forecasts, which impose the cointegration constraint, are superior to those from an unrestricted vector autoregressive model although this is not the case in short- to medium-term forecasts. Lin and Tsay (1996) also showed experimentally that correct specification of the cointegrating relationship leads to better long-term forecasts, although this is again not the case for short- to medium-term forecasts. Reinsel and Ahn (1992) obtained experimental results which indicate that imposing the cointegration constraint helps to improve long-term forecasts. These earlier studies have spread the belief that if the specification of the cointegrating relationship is correct, imposing cointegration helps long-term forecasts.

Many of the previous studies have been concerned with small cointegrated processes, i.e., processes in which the number of variables in the process (hereafter, m) is typically less than 5. Here, we consider forecasting in large cointegrated processes in which m is greater than 20. An example would be the forecasting of the stock prices of 30 companies in an industry. The stock price of each company is generally considered to be well approximated by a random walk or an ARIMA process. However, companies in the same industry face the same market conditions and macroeconomic environment. Thus, their stock prices tend to move together, and it appears that they constitute a large cointegrated process.

There remain a few practical problems in forecasting large cointegrated processes. First, it is difficult to determine the cointegration rank (hereafter, r) in large processes. The widely used Johansen (1988, 1991) procedure is applicable only for processes with $m = 10$ or less, since the critical values for the tests are available up to 10 variables.(See, for example, the tables in Osterwald-Lenum (1992).) This means that in large models, the cointegration relationship is hard to specify correctly. Second, in the conventional forecasting procedure with vector error correction (hereafter, VEC) models, the number of parameters to be estimated rapidly increase with m. When the number of parameters to be estimated is large relative to the sample size, estimators will have large standard errors, and consequently forecasts will also have large standard errors.

The purpose of the present paper is to propose a new forecasting procedure that has two new features in order to avoid the above difficulties. First, in order to determine r, we adopt the test procedure for the cointegration rank recently developed by Chigira (2006), which is a slight generalization of Harris (1997) and Snell (1999) and can be applied to any large cointegrated process. Second, in order to economize the number of parameters, we resort to the forecasting procedure by Kariya (1988). Further, we generalize Kariya's procedure so that our forecast imposes the cointegration constraint. We note that our forecasting procedure is also robust in comparison with the conventional procedure with VEC models since it neither requires a VAR representation of the process nor Gaussian innovations. In this paper, we conduct a Monte Carlo simulation to evaluate the effect of imposing cointegration on our forecast.

However, before presenting our experimental results, it is useful to critically discuss the effect of imposing cointegration on forecast accuracy. As mentioned above, Engle and Yoo (1987), Lin and Tsay (1996), and Reinsel and Ahn (1992) argued that while imposing cointegration does not improve short- to medium-term forecasts, it does improve long-term forecasts. Their argument is based on their Monte Carlo experiments which compared the accuracy of forecasts which impose both integration and cointegration to that of forecasts which impose neither. However, as Christoffersen and Diebold (1998) noted, such a comparison does not necessarily show the effect of imposing cointegration. Instead, they argued, a comparison between forecasts which impose both integration and cointegration and forecasts which impose only integration would be appropriate. They analytically showed that in the long-run, Engle and Yoo's (1987) cointegrated system forecast (which imposes both integration and cointegration) performs as well as a univariate ARIMA forecast of an individual series (which imposes only integration) in terms of the mean square error (MSE) criterion. That is, imposing cointegration does not matter in long-term forecasts.[1] On the other hand, Christoffersen and Diebold (1998)

[1] Chigira and Yamamoto (2006) extended Christoffersen and Diebold's (1998) result and analytically showed that in terms of the MSE criterion, neither integration nor cointegration matters in long-term forecasts.

noted, imposing cointegration does help short- to medium- term forecasts. Their result, i.e., that imposing cointegration helps to improve not long-term forecasts but short- to medium- term forecasts, is the exact opposite of the results obtained by Engle and Yoo, etc.

Peña and Poncela (2004) remarked on the result of Christoffersen and Diebold (1998). Peña and Poncela (2004) analyzed the forecasts of the dynamic factor model[2] and obtained results which indicate that imposing cointegration does help long-term forecast. Explicitly, they analytically showed that in one factor models, the factor model forecasts outperform the univariate ARIMA forecasts in the long-run and the advantage of the factor model forecasts increases with m. Peña and Poncela (2004) argued that one reason why Christoffersen and Diebold found that imposing cointegration does not matter could be that they did not consider the effect of the number of variables.

Our Monte Carlo experiments, which evaluate the effect of imposing cointegration on our forecast, are designed to respond to these results. Explicitly, we compare the accuracy of our forecast to that of the univariate ARIMA forecast of an individual series, as suggested by Christoffersen and Diebold (1998). Further, we examine whether imposing cointegration improves the accuracy of forecasts in short-run, say up to 5 periods ahead or in medium-run, say 10 to 30 periods ahead, rather than in the long-run. This is because Christoffersen and Diebold (1998) analytically showed that imposing cointegration helps not the long-term forecast but short to medium term forecasts. In addition, taking the remark of Peña and Poncela (2004) into consideration, we investigate whether m affects the forecast accuracy.

In the Monte Carlo simulation we conduct below, we find that our forecast outperforms the univariate ARIMA forecast. This result stems from the fact that our forecast imposes cointegration. We also find that our forecast is superior to conventional forecasts based on the VEC model, especially in small samples, which is due to the fact that our forecasting procedure economizes on the number of parameters. Our Monte Carlo results show that as m becomes large, as long as the ratio r/m is sufficiently high, the advantage

[2] It is well known that cointegrated time series models are closely related to dynamic factor models whose factors are nonstationary. See, e.g., Escribano and Peña (1994).

of our forecast over the univariate ARIMA forecast increases. This result is similar to the result of Peña and Poncela (2004) who argued that in one factor models, the advantage of the factor model forecasts over the univariate ARIMA forecast increases with m. This is an expected result because cointegrated time series models whose r/m -ratio is high are closely related to one factor models. Specifically, because one factor models indicate that the number of common stochastic trends of the models is one, i.e., $m - r = 1$, one factor models are closely related to cointegrated time series models whose r/m -ratio is high: $r/m = (m-1)/m \approx 1$. In addition, an interesting result of our analysis is that as r becomes large relative to m, the accuracy of our forecast improves. Intuitively, this is because our forecasting procedure explicitly takes into account the information on the cointegration rank.

The remainder of this paper is organized as follows. Section 2 describes our model and reviews the accuracy measure of forecasts. Section 3 explains the new testing procedure for the cointegration rank developed by Chigira (2006) which can be applied to large systems. Section 4 develops our forecasting procedure, which is economical with the number of parameters. Next, Section 5 reports the results of our Monte Carlo experiments, while Section 6 contains an empirical illustration of stock price forecasts for 25 Japanese pharmaceutical companies. Section 7 offers some concluding remarks.

2 THE MODEL AND THE ACCURACY MEASURE

In this paper we consider an m-dimensional vector cointegrated process which can be represented as a vector moving average of infinite-order

$$(1-L)y_t = \mu + C(L)\varepsilon_t = \mu + \sum_{i=0}^{\infty} C_i \varepsilon_{t-i} \tag{1}$$

where L is the lag operator, μ is a vector of constants, C_i is an $m \times m$ matrix with absolute summability of $\{sC_s\}_{s=0}^{\infty}$, ε_t is an $m \times 1$ i.i.d.$(0, \Omega)$ process. Under regularity conditions, a necessary and sufficient condition for the existence of an $m \times r$ matrix of cointegration β such that $z_t = \beta' y_t$ follows a zero-mean $I(0)$ process, is given by $\beta' C(1) = 0$ with $\text{rank}(C(1)) = m - r$.

We now explain the accuracy measure used in this paper, the MSE. The MSE is a standard accuracy measures for forecasting in scalar processes. In m-variate processes, the MSE is $E(e'_{t+h}Ke_{t+h})$ where e_{t+h} is an h-period-ahead forecast error vector and K is an $m \times m$ positive (semi-)definite symmetric matrix. Usually, the weight matrix K is set to be I_m, i.e., the trace($E(e_{t+h}e'_{t+h})$) is often used as an accuracy measure. They argue that in order to investigate the relative accuracy of two forecasts, say 1 to 2, it is standard to use the ratio

$$\frac{\text{trace MSE}(e_{1,t+h})}{\text{trace MSE}(e_{2,t+h})}. \tag{2}$$

Following Christoffersen and Diebold (1998), we call this the "trace MSE ratio." In this paper we adopt the trace MSE ratio as the measure of accuracy. As Christoffersen and Diebold (1998) and Clements and Hendry (1993, 1995) stressed, the accuracy measure should be selected to be consistent with forecasters' loss function. We believe that many forecasters are interested in the trace MSE ratio because it provides an easily understandable summary of the forecast errors. In addition, it has been widely used in previous studies including Engle and Yoo (1987) and Lin and Tsay (1996).[3]

Using the MSE ratio as an accuracy measure, Christoffersen and Diebold (1998) investigated whether imposing cointegration improves the accuracy of forecasts. Specifically, they examined the relative accuracy of Engle and Yoo's (1987) cointegrated system forecast (defined below) and the univariate ARIMA forecast, which is an ARIMA(p,1,q) forecast of each element of y_t. The cointegrated system forecast for period $t+h$ is represented as

$$\hat{y}_{t+h} = (t+h)\mu + \sum_{i=1}^{t} \sum_{j=0}^{t+h-i} C_j \varepsilon_i, \tag{3}$$

and its forecast error is given by

$$\hat{e}_{t+h} = y_{t+h} - \hat{y}_{t+h} = \sum_{i=1}^{h} \sum_{j=0}^{h-i} C_j \varepsilon_{t+i}. \tag{4}$$

Engle and Yoo (1987) showed that forecast $\hat{y}_{t+h}$ imposes the cointegration constraint

$$\lim_{h\to\infty} \beta' \hat{y}_{t+h} = 0. \tag{5}$$

[3] See Lin and Tsay (1996) for a more detailed discussion of reasons to choose the trace MSE ratio.

In contrast, the univariate ARIMA forecast, while it imposes integration, does not satisfy the cointegration constraint because it is based on an ARIMA(p,1,q) model of an individual series. Christoffersen and Diebold (1998) argued that because the cointegrated system forecast imposes both integration and cointegration and the univariate ARIMA forecast imposes only integration, the trace MSE ratio for these two forecasts allows us to evaluate the effect of imposing cointegration on the forecast accuracy. They analytically showed that

$$\lim_{h\to\infty} \frac{\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))} = 1, \tag{6}$$

where $\tilde{y}_{t+h}$ denotes the univariate ARIMA forecast and $\tilde{e}_{t+h} = y_{t+h} - \tilde{y}_{t+h}$. Equation (6) indicates that under the trace MSE ratio, imposing cointegration does not matter in long-term forecasts.[4] Christoffersen and Diebold (1998) also found that the cointegrated system forecast is superior to the univariate ARIMA forecast in the short- to medium-term. As mentioned in the introduction, this result is the opposite of that of Engle and Yoo (1987) and others who argued that imposing cointegration does not matter in short- to medium-term forecasts but does so in long-term forecasts.

In order to evaluate the effect of imposing cointegration on our forecast, we take the above results into consideration and in Section 5 calculate the trace MSE ratio of the forecast error of our forecast and of the univariate ARIMA forecast.

3 TEST FOR COINTEGRATION

In order to impose cointegration on forecasts, it is necessary to determine the cointegration rank. However, as noted in the introduction, testing for the cointegration rank in processes with more than 10 variables is difficult. In this section, we introduce a new cointegration rank test developed by Chigira (2006) that is independent of the number of variables in the process.

[4] Christoffersen and Diebold (1998) noted that this result depends on the adopted accuracy measure. They proposed alternative accuracy measures under which the cointegrated system forecast outperforms the univariate ARIMA forecast in the long-run.

Proposition 1 (Chigira (2006)): Let $\{y_t\}_{t=1}^T$ *be an observed* m*-variate time series that can be represented as*

$$y_t = \mu + \delta t + x_t, \tag{7}$$

where μ *and* δt *are the deterministic components and* x_t *is an* $I(1)$ *and cointegrated system satisfying the following equations:*

$$\beta_\perp' x_t \sim I(1), \tag{8}$$

$$\beta' x_t = O_p(1), \tag{9}$$

where β *is an* $m \times r$ *matrix of* r *cointegrating vectors with rank* r*,* $\beta_\perp$ *is of full rank and dimension* $m \times (m-r)$ *such that* $\beta'\beta_\perp = 0$ *and*

$$\beta' S_{\bar{x}\bar{x}} \beta \xrightarrow{p} \text{positive definite matrix}, \tag{10}$$

$$\beta_\perp' S_{\bar{x}\bar{x}} \beta_\perp = O_p(T), \tag{11}$$

$$\beta_\perp' S_{\bar{x}\bar{x}} \beta = O_p(1), \tag{12}$$

where $S_{\bar{x}\bar{x}} = \frac{1}{T}\sum_{t=1}^T \bar{x}_t \bar{x}_t'$ *with* $\bar{x}_t$ *is the series after removing the mean value and linear trend from* y_t*. Then, if* T *is sufficiently large, the principal components* $B'\bar{x}_t$ *of a series* $\bar{x}_t$ *have the following properties:*

(i) $B_{(m-r)}' \bar{x}_t \sim I(1)$,

(ii) $B_{(r)}' \bar{x}_t = O_p(1)$,

where $B = [B_{(m-r)},\ B_{(r)}]$ *is an* $m \times m$ *matrix of eigenvectors such that*

$$S_{\bar{x}\bar{x}} = B \Lambda B', \tag{13}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 & \\ & \lambda_2 & & \\ & & \ddots & \\ & 0 & & \lambda_m \end{bmatrix}, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m, \tag{14}$$

where $B_{(m-r)}$ *is the* $m \times (m-r)$ *matrix of eigenvectors corresponding to the largest* $m-r$ *eigenvalues and* $B_{(r)}$ *is the* $m \times r$ *matrix of eigenvectors corresponding to the smallest* r *eigenvalues.*

Proof: See Chigira (2006).

Remark 1: The above proposition is a slight generalization of Harris (1997) and Snell (1999). The basic difference lies in the fact that assumptions (10) - (12) of the proposition are weaker than their assumptions. Explicitly, Harris (1997) assumes that $\beta' x_t$ and $\beta'_\perp(1-L)x_t$ are stationary and satisfy certain regularity conditions and Snell (1999) assumes that $\beta' x_t$ follows an AR(p) process and $\beta'_\perp(1-L)x_t$ follow an MA(∞) process.

Remark 2: We note that our model (1) can be represented in the form of (7) and satisfies (8) and (9). In addition, Park and Phillips (1988) show that (10) - (12) hold for our model (1). Thus, we can apply the result of the above proposition in this paper.

The above proposition allows us to construct a sequential testing procedure to decide the cointegration rank. For example, we first test the null of $r = 0$, that is to say, apply a unit root test for the mth principal component. If the null hypothesis is accepted, the cointegration rank is decided to be zero. If rejected, we then test the null hypothesis of $r = 1$ or apply a unit root test for the $(m-1)$th principal component. We sequentially continue to test the null of a unit root for the $(m-r)$th principal component until it is accepted. When the null of a unit root is accepted for the $(m-r)$th principal component, the cointegration rank is determined to be r. Obviously, this procedure is applicable regardless of the number of variables in the process. The Monte Carlo experiments in Chigira (2006) show that this sequential procedure works well even if the sample size is small and/or the disturbance is serially correlated. The Monte Carlo experiments indicate that the test also performs well in a large cointegrated system.

In order to determine the degree of integration of each principal component, we apply the Phillips and Perron (hereafter, PP) type unit root test (Phillips (1987) and Phillips and Perron (1988)). Thus, we do not have to assume a VAR representation of the process or normal error disturbances, which indicates that the present procedure is more robust than Johansen's (1988, 1991) procedure.

Of course, Chigira (2006) is not the only test of cointegration in large cointegrated processes. Bai and Ng (2004), for example, proposed various tests in dynamic factor models which appear to be applicable to large cointegrated processes. This is because, as mentioned in the introduction, cointegrated time series models and dynamic factor models are closely related. However, Bai and Ng's (2004) approach, which is based on large m asymptotics, is different from our approach in which m is large but finite and fixed. In addition, the critical values of their cointegration tests depend on r under the null hypothesis. Thus, their tests are not suitable for our approach.

4 METHODS FOR FORECASTING IN LARGE COINTEGRATED SYSTEMS

In this section, we begin by presenting our forecasting method. We then review the conventional forecasting method based on the VEC model, which is a feasible version of the cointegrated system forecast.

4.1 The MTV Forecast

As mentioned earlier, the conventional forecasting procedure with the VEC model requires the estimation of a large number of parameters. Specifically, the number of coefficient parameters to be estimated with the VEC model is $O(m^2)$ (see Section 4.2). Thus, in large processes, the forecasting procedure based on the VEC model tends to give forecasts whose variances of forecast errors are large for the case of small samples. In order to economize on the number of parameters to be estimated, we here develop a forecasting method that modifies Kariya's (1988) approach.[5]

The basic idea of our forecasting method is to estimate and forecast each principal component of y_t as a univariate ARIMA(p,d,q) process instead of estimating and forecasting the m-variate process simultaneously. We then combine the forecasts from individual ARIMA models so as to satisfy the cointegration constraint. Following Kariya's (1988)

[5] There are a few fundamental differences between Kariya's (1988) original method and ours. First, he assumes stationarity of the process concerned and consequently does not pay attention to integration or cointegration. Second, he follows the conventional 95% rule of principal component analysis and throws away principal components which correspond to smaller characteristic roots contributing less than 5% of total variation. In contrast, we use all principal components in forecasting. Experiments we conducted but which are not reported here showed that throwing away those principal components results in a substantial loss of accuracy in forecasting.

original terminology, we refer to this approach as the multivariate time series variance component (MTV) approach.

We first present the procedure of the MTV approach and then provide a few remarks on it.

The MTV Procedure:

Step 1: Calculate B in (13) and apply a unit root test sequentially to the elements of $B'\bar{y}_t$, where $\bar{y}_t$ is the series after removing the mean value and linear trend from y_t.

Step 2: Make an h-period-ahead forecast from a suitably fitted univariate ARIMA$(p,1,q)$ model when an element is judged to contain a unit root. Make an h-period-ahead forecast from a suitably fitted univariate ARMA(p,q) model when an element is judged to be stationary. Stack these forecasts and denote the vector as $\widehat{B'\bar{y}}_{t+h}$.

Step 3: Make the MTV forecast for the original series, denoted $\hat{\hat{y}}_{t+h}$, by multiplying B and $\widehat{B'\bar{y}}_{t+h}$ and adding the constant drift so that $\hat{\hat{y}}_{t+h} = B\widehat{B'\bar{y}}_{t+h} + (t+h)\hat{\mu}$, where $\hat{\mu}$ is a consistent estimator of μ.

The MTV forecast for model (1) can be written as:

$$\begin{aligned}
\hat{\hat{y}}_{t+h} &= (t+h)\hat{\mu} + B\widehat{B'\bar{y}}_{t+h} \\
&= (t+h)\hat{\mu} + B_{(m-r)}\widehat{B'_{(m-r)}\bar{y}}_{t+h} + B_{(r)}\widehat{B'_{(r)}\bar{y}}_{t+h} \\
&\approx (t+h)\mu + \beta_\perp Q_\perp \widehat{B'_{(m-r)}\bar{y}}_{t+h} + \beta Q \widehat{B'_{(r)}\bar{y}}_{t+h} \qquad (15)
\end{aligned}$$

where $B_{(m-r)}$ and $B_{(r)}$ are given in Proposition 1, $Q_\perp = (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp B_{(m-r)}$, and $Q = (\beta'\beta)^{-1}\beta' B_{(r)}$. The approximation in line (15) holds as T becomes large, since, as shown by Chigira (2006):

$$B_{(m-r)} = \beta_\perp Q_\perp + O_p(T^{-1}), \qquad (16)$$

$$B_{(r)} = \beta Q + O_p(T^{-1}). \qquad (17)$$

Remark 4: Harris (1997) and Snell (1999) also derive (16) and (17). However, their underlying assumptions are more restrictive. It may be noted that Chigira (2006) derives

(16) and (17) based on assumptions (10) - (12), which are less restrictive than those of Harris (1997) and Snell (1999).

We now show some of the properties of the MTV forecast.

Proposition 2: If T is sufficiently large,

(i) *the MTV forecast imposes the cointegration constraint:*

$$\lim_{h\to\infty} \beta' \hat{\hat{y}}_{t+h} = 0.$$

(ii) *in terms of our measure of accuracy, i.e., the trace MSE ratio, the MTV forecast performs as well as the cointegrated system forecast in the long-run:*

$$\lim_{h\to\infty} \frac{\text{trace}(E(\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}))}{\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))} = 1,$$

where $\hat{\hat{e}}_{t+h} = y_{t+h} - \hat{\hat{y}}_{t+h}$.

(iii) *in terms of our measure of accuracy, i.e., the trace MSE ratio, the MTV forecast performs as well as the univariate ARIMA forecast in the long-run:*

$$\lim_{h\to\infty} \frac{\text{trace}(E(\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))} = 1.$$

(iv) *in terms of the alternative measure of accuracy, the trace MSE ratio with the weight matrix K being $\beta\beta'$,*[6] *the MTV forecast performs as well as the cointegrated system forecast and outperforms the univariate ARIMA forecast in the long-run:*

$$\lim_{h\to\infty} \frac{\text{trace}(E(\beta'\hat{e}_{t+h}\hat{e}'_{t+h}\beta))}{\text{trace}(E(\beta'\tilde{e}_{t+h}\tilde{e}'_{t+h}\beta))} = \lim_{h\to\infty} \frac{\text{trace}(E(\beta'\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}\beta))}{\text{trace}(E(\beta'\tilde{e}_{t+h}\tilde{e}'_{t+h}\beta))} < 1.$$

Proof: See Appendix.

Proposition 2(i) shows that the long-term MTV forecast imposes cointegration. Note, however, that in practice, it is impossible to impose cointegration on the MTV forecast without knowing r. The reason is that if we do not know r, then we cannot divide B into $B_{(r)}$ and $B_{(m-r)}$, as in equation (15). Thus, we need to determine the cointegration rank.

[6] This is the measure proposed by Christoffersen and Diebold (1998) under which the cointegrated system forecast outperforms the univariate ARIMA forecast in the long-run.

Fortunately, going through Step 1 of the MTV procedure, for the reasons mentioned in Section 3, allows us to determine the cointegration rank, which enables us to impose cointegration on the MTV forecast. Therefore, in practice, we can impose cointegration on the MTV forecast for any large cointegrated process.

We can see Proposition 2(iii) as the MTV version of equation (6). As in the case of the cointegrated system forecast, the trace MSE ratio suggests that the MTV forecast and the univariate ARIMA forecast are equally poor in the long-run. Note, however, that the MTV forecast imposes cointegration, as shown in Proposition 2(i). Because Christoffersen and Diebold (1998) argued that imposing cointegration helps to improve the cointegrated system forecast in the short- to medium-run, we expect that the MTV forecast outperforms the univariate ARIMA forecast (which does not impose cointegration) in the short- to medium-run. However, for finite h, it is difficult to calculate analytically the trace MSE ratio of the MTV forecast and the univariate ARIMA forecast. Also, analytically, it is difficult to determine whether, for finite h, the MTV forecast outperforms the cointegrated system forecast, although Proposition 2(ii) analytically shows that they are equally performed as h goes to infinity. In Section 5, therefore, we carry out Monte Carlo experiments to examine these issues experimentally.

4.2 The VEC Forecast

In practice, the cointegrated system forecast is hard to calculate because it is difficult to estimate $\{C_i\}_{i=1}^{\infty}$. In order to make the cointegrated system forecast feasible, it is conventional to assume that the m-variate process of interest can be represented as a vector autoregressive model. For ease of exposition, we consider a VAR model without deterministic terms:

$$A(L)y_t = \varepsilon_t, \tag{18}$$

where $A(L)$ is a lag polynomial such that $A(L) = I_m - A_1 L - \cdots - A_p L^p$ and ϵ_t is distributed $i.i.d.N(0, \Omega)$. Under regularity conditions, it is ensured that each element of y_t is $I(1)$ and the matrix β consists of cointegrating vectors in the sense that $\beta' y_t$ is stationary. Reparameterizing (18), we find the VEC representation of the model:

$$(1-L)y_t = \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \tag{19}$$

where $\Gamma_j \equiv -\sum_{i=j+1}^{p} A_i$ for $j = 1, \cdots, p-1$. From (18), we find that the number of parameters to be estimated is $m^2(p-1)+mr$. For forecasting, it is convenient to rewrite equation (18) as

$$Y_t = \bar{A} Y_{t-1} + \Xi_t, \tag{20}$$

where $Y_t' = [y_t', y_{t-1}', \cdots, y_{t-p}']$, $\Xi_t' = [\epsilon_t', 0, \cdots, 0]$,

$$\bar{A} = \left[\begin{array}{c} A \\ \hline I_{(p-1)m} \mid 0 \end{array}\right] = \begin{bmatrix} A_1 & A_2 & \cdot & \cdot & \cdot & A_{p-1} & A_p \\ I_m & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & I_m & & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & & \cdot & \cdot & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & I_m & 0 \end{bmatrix},$$

$A_1 = I_m + \alpha\beta' + \Gamma_1$ and $A_i = \Gamma_i - \Gamma_{i-1}$ for $i = 2, \cdots, p-1$. Using equation (20), the VEC forecast for period $t+h$ is calculated as

$$\hat{y}_{t+h} = M' \bar{A}^h Y_t, \tag{21}$$

where $M' = [I_m, 0, \cdots, 0]$.

The conventional procedure is summarized as follows:

The VEC Procedure:

Step 1: Assuming that the process is represented by the VEC form, determine the cointegration rank r using the well-known Johansen (1988, 1991) tests.

Step 2: Given r, obtain maximum likelihood estimates $\hat{\beta}$, $\hat{\alpha}$ and $\hat{\Gamma}_1, \cdots, \hat{\Gamma}_{p-1}$.

Step 3: Recover $\hat{A}_1, \cdots, \hat{A}_p$ from $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\Gamma}_1, \cdots, \hat{\Gamma}_{p-1}$ and calculate forecast $\hat{y}_{t+h}$ by (21).

Remark 3: Note that it is difficult to determine r using the Johansen tests in large processes. Therefore, in Step 1 of the VEC procedure, we may use a test that is easy to calculate and independent of m rather than the Johansen tests. In Sections 5 and 6, we will use the test proposed in Section 3 instead of the Johansen tests.

As long as the process of interest follows model (19) and satisfies regularity conditions, the VEC forecast is equal to the cointegrated system forecast. Even if some assumptions of the VEC approach are violated, we can approximate the cointegrated system forecast by the VEC forecast. For example, when the independence of the error term of model (19) is violated, the cointegrated system forecast is often approximated by the VEC forecast based on the lag-augmented VEC model suggested by Said and Dickey (1984). Because in practice parameters C_i must be estimated, in the following section, we choose the VEC forecast rather than the cointegrated system forecast as a competitor for the MTV forecast.

5 MONTE CARLO EXPERIMENT

In this section, we present the results of our Monte Carlo experiments. The purpose is to investigate whether the MTV forecast outperforms the VEC forecast and the univariate ARIMA forecast in the short- to medium-run in terms of the trace MSE ratio.

However, before presenting the Monte Carlo experiments, we briefly highlight the merits of the MTV approach. In comparison with the VEC approach, the MTV approach has the following advantages: First, it assumes model (1) rather than (18) and the error term is not necessarily normally distributed. Thus, it is less restrictive than the conventional VEC approach. Second, the MTV approach economizes the number of parameters to be estimated. On the other hand, in comparison with the univariate ARIMA forecast, the MTV forecast has the important advantage that it imposes cointegration.

5.1 Design of the Experiment

In our experiment, we consider three data generating processes (DGPs). They are given by the following models:

$$\text{Case 1: } (1-L)y_t = \alpha\beta' y_{t-1} + \varepsilon_t, \tag{22}$$

$$\text{Case 2: } (1-L)y_t = \alpha\beta' y_{t-1} + u_t,\ \ u_t = 0.5\varepsilon_{t-1} + \varepsilon_t, \tag{23}$$

$$\text{Case 3: } (1-L)y_t = \alpha\beta^{+\prime} y_{t-1}^{+} + u_t,\ \ u_t = \Psi\nu_{t-1} + \nu_t,\ \ \nu_t = \Omega^{1/2}\varepsilon_t, \tag{24}$$

where α and β are the loading matrix and the cointegration matrix, respectively, $y_{t-1}^{+} = [1, y_{t-1}']'$, $\beta^{+} = [\rho_0, \beta']'$ with ρ_0 being an $r \times 1$ parameter vector, Ψ is an $m \times m$ parameter matrix whose characteristic roots are less than unity in absolute value, Ω is an $m \times m$ positive definite matrix, and ε_t is an $m \times 1$ random vector distributed $i.i.d.N(0,1)$. We set α, β, ρ_0, Ψ and Ω using a random number generator but do not report their values due to space limitations.

We consider the following different configurations of m, r, h, and T in the experiment:

(i) The number of variables in the system m: m =5 and 25.

(ii) The cointegration rank r for a given m: r =2 and 4 for m =5, and r =10 and 20 for m =25.

(iii) Prediction horizon h: h =5, 10, 15, and 30.

(iv) Sample size T: T =100, 500, and 1000.

Note that in Case 1, the cointegrated system forecast is equal to the VEC forecast. In other words, the VEC approach is justified. In Cases 2 and 3, we approximate the cointegrated system forecast by the VEC forecast, following the suggestion of Said and Dickey (1984). Thus, we consider the VEC forecast rather than the cointegrated system forecast as a competitor for the MTV forecast. When we calculate the VEC forecast, we follow the VEC Procedure. As noted in Remark 3, instead of the Johansen tests, we use the testing procedure described in Section 3 to determine the cointegration rank, i.e., we apply the PP test for a unit root to the principal components, using a 1% significance level. In the VEC approach, the lag length $p = 1$ is treated as known. To calculate the MTV forecast, we follow the MTV Procedure. In fitting a univariate ARIMA$(p, 1, q)$ or ARMA(p, q) model to each principal component, p and q are determined by the Schwartz Bayesian Information Criterion (SBIC). We calculate the univariate ARIMA forecast by fitting each element of y_t to a univariate ARIMA$(p, 1, q)$ model, with p and q being determined by the SBIC.

In order to compare the accuracy of these three forecasts, we calculate the following trace MSE ratios with 5000 replications:

$$\text{tr MSE(VEC)} = \frac{\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))} \text{ and}$$

$$\text{tr MSE(MTV)} = \frac{\text{trace}(E(\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))}.$$

5.2 Results for Case 1: IID Normal Errors

Case 1 represents a situation where the VEC approach is justified. The results are given in Table 1.

The MTV forecast versus the univariate ARIMA forecast

In Table 1, "tr MSE(MTV)" reports the result of the comparison between the MTV forecast and the univariate ARIMA forecast. We find that the values of tr MSE(MTV) are mostly less than 1, which means that the MTV forecast almost always outperforms the univariate ARIMA forecast. The advantage of the MTV forecast over the univariate ARIMA is attributable to imposition of the cointegration. However, the MTV forecast loses some of its advantage over the univariate ARIMA forecast when $m = 25$ and $T = 100$.

The MTV forecast versus the VEC forecast

Comparing tr MSE(MTV) with tr MSE(VEC) in Table 1, we can determine whether the MTV forecast outperforms the VEC forecast. In the case of $m = 5$, the values of tr

Table 1: Performance of the VEC and the MTV Approach - Case 1

m =5

T	r	tr MSE(VEC)					tr MSE(MTV)				
		h					h				
		1	5	10	15	30	1	5	10	15	30
100	2	0.99	0.98	0.98	0.99	1	1.01	0.98	0.98	0.98	0.99
	4	0.99	0.94	0.94	0.94	0.96	0.98	0.91	0.9	0.91	0.94
500	2	0.95	0.89	0.92	0.94	0.98	0.94	0.89	0.92	0.94	0.97
	4	0.95	0.85	0.85	0.88	0.93	0.95	0.85	0.85	0.88	0.92
1000	2	0.95	0.9	0.92	0.94	0.97	0.95	0.9	0.92	0.94	0.97
	4	0.95	0.87	0.87	0.89	0.93	0.95	0.87	0.87	0.89	0.93

m =25

T	r	tr MSE(VEC)					tr MSE(MTV)				
		h					h				
		1	5	10	15	30	1	5	10	15	30
100	10	1.36	1.29	1.21	1.17	1.1	1.15	1.07	1.03	1.02	1.01
	20	1.3	1.14	1.08	1.07	1.04	1.09	0.97	0.96	0.96	0.97
500	10	0.99	0.98	1	1	1.01	0.98	0.96	0.97	0.98	0.99
	20	0.97	0.86	0.84	0.87	0.92	0.94	0.83	0.83	0.86	0.91
1000	10	0.96	0.92	0.95	0.96	0.98	0.96	0.92	0.94	0.96	0.98
	20	0.96	0.85	0.85	0.87	0.92	0.94	0.83	0.84	0.86	0.91

Note: tr MSE(VEC) $= \frac{\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))}$ where $\hat{e}_{t+h}$ is the forecast error from the VEC forecast and $\tilde{e}_{t+h}$ is the forecast error from the univariate ARIMA forecast. tr MSE(MTV) $= \frac{\text{trace}(E(\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))}$ where $\hat{\hat{e}}_{t+h}$ is the forecast error from the MTV forecast.

MSE(MTV) are mostly equal to those of tr MSE(VEC), i.e., the MTV forecast performs as well as the VEC forecast. In contrast, in the case of $m = 25$, most tr MSE(MTV) values are lower than the tr MSE(VEC) values, i.e., the MTV forecast outperforms the VEC forecast. For example, when $m = 25$, $T = 100$, $r = 20$ and $h = 5$, tr MSE(MTV)=0.97 and tr MSE(VEC)=1.14. A possible reason for the advantage of the MTV forecast over the VEC forecast is that it economizes on the number of parameters to be estimated. For $T = 500$ and 1000, tr MSE(VEC) becomes closer to tr MSE(MTV). This result indicates that the VEC forecast needs at least $T = 500$ to work as well as the MTV forecast.

Increasing m

Taking into account the remarks of Peña and Poncela (2004) mentioned in the introduction, we investigate whether m affects the relative accuracy of the forecasts. Table 1 shows that the MTV forecast almost always outperforms the univariate ARIMA forecast in the case of both $m = 5$ and 25. Table 1 also shows that the relative advantage of the MTV forecast over the univariate ARIMA forecast becomes more evident when $m = 25$ than when $m = 5$ as long as T is large and r/m is close to unity. For example, with $T = 1000$, $r/m = 0.8$ and $h = 5$, tr MSE(MTV)=0.83 when $m = 25$ and 0.87 when $m = 5$. This result appears to be consistent with Peña and Poncela (2004), who showed that in one factor models, i.e., where $r/m = (m-1)/m \approx 1$, the advantage of the dynamic factor model forecast over a univariate ARIMA forecast becomes more evident as m increases. The VEC forecast exhibits a similar tendency, i.e., tr MSE(VEC) becomes smaller as m increases as long as T is large and r/m is high.

Increasing r

An interesting result in Table 1 is that as r becomes large relative to m, the accuracy of the MTV forecast relative to the univariate ARIMA forecast improves. For example, with $m = 25$, $T = 500$ and $h = 5$, tr MSE(MTV)=0.96 when $r = 10$ and 0.83 when $r = 20$. The VEC forecast exhibits a similar tendency. An intuitive explanation of this result is that because the VEC and MTV approaches explicitly use the information on the cointegration rank, the number of unit roots in the VEC and MTV forecasts is only $m-r$. In contrast, the number of unit roots in the univariate ARIMA forecast invariably is m, irrespective of the cointegration rank.

Increasing h

Overall, tr MSE(VEC) and tr MSE(MTV) reach each minimum each at $h = 5$ or 10. This means that imposing cointegration improves the VEC and the MTV forecast in the short- to medium-run. On the other hand, as h becomes larger than 10, both tr MSE(VEC) and tr MSE(MTV) get large and closer to 1. This result indicates that in the long-run, imposing cointegration does not matter in terms of the trace MSE ratio, which is consistent with equation (6) and Proposition 2(iii).

5.3 Results for Case 2: MA(1) Errors

Case 2 represents a situation where the error term exhibits a serial correlation. In order to apply the VEC approach in an appropriate manner, we approximate model (23) by a VEC model with a lag order of $p = 1$ when $T = 100$, 2 when $T = 500$, and 3 when $T = 1000$, following the suggestion of Said and Dickey (1984). The results are given in Table 2. Comparing Table 2 with Table 1, we examine whether the MTV forecast and the VEC forecast are robust to a serial correlation in the error term.

As can be seen from Table 2, the results for tr MSE(MTV) are quite similar to those in Table 1. This result indicates that the MTV forecast is robust to a serial correlation in the error term. In fact, when we provide the proof for Proposition 2(i) in Appendix, we do not use the assumption of serial independency for the error term. Therefore, this result is as expected.

In contrast to tr MSE(MTV), the values of tr MSE(VEC) are uniformly greater than those in Table 1. Unsurprisingly, given that the VEC approach depends on the assumption of serial independence of the error term, the accuracy of the VEC forecast becomes low. In particular, when the sample size is small and $m = 25$, the values of tr MSE(VEC) are quite large. In Case 2, the MTV forecast clearly outperforms the VEC forecast.

The results of a detailed analysis of tr MSE(MTV) for various combinations of m, r, h, and T are quite similar to those of Case 1 and therefore omitted.

5.4 Results for Case 3: A Deterministic Term and MA(1) Errors with Heteroscedasticity

The DGP in Case 3 has an error term which exhibits a serial correlation. Furthermore, in order to conform with practical situations, the DGP also contains a deterministic term and exhibits contemporaneous correlation caused by Ψ and Ω. In our experiment, we calculate the VEC forecasts based on the VEC model which correctly specifies the deterministic term. As in Case 2, we augment the lag length of the VEC model.

Table 3 reports the results of the experiment. The overall pattern of the MTV forecast is similar to that found in Cases 1 and 2, although tr MSE(MTV) is uniformly larger than in Cases 1 and 2. In the case of $m = 25$ and $T = 100$, the univariate ARIMA forecast outperforms the MTV forecast. However, as long as T is large, the MTV forecast almost always outperforms the univariate ARIMA forecast, which indicates that the MTV forecast is robust to heteroscedasticity and contemporaneous correlation in the error term. On the other hand, the performance of the VEC forecast becomes seriously deteriorates. In Case 3, just like in Case 2, the MTV forecast works better than the VEC forecast.

5.5 Summary of the Experiment

The results of the experiment can be summarized as follows:

(i) The MTV forecast almost always outperforms the univariate ARIMA forecast, except for the case of $T = 100$. This result is the consequence of the fact that the MTV forecast imposes cointegration.

(ii) The MTV forecast almost always outperforms the VEC forecast. It is interesting to note that this is true even in Case 1 where the VEC approach is justified, although the difference between tr MSE(VEC) and tr MSE(MTV) is small. This may suggest

Table 2: Performance of the VEC and the MTV Approach - Case 2

m =5

T	r	tr MSE(VEC)					tr MSE(MTV)				
		h					h				
		1	5	10	15	30	1	5	10	15	30
100	2	1.17	1	1	1.01	1.01	1	0.98	0.98	0.99	0.99
	4	1.12	1.01	1.01	1.01	1.01	0.99	0.97	0.96	0.97	0.98
500	2	1.01	0.92	0.93	0.95	0.98	0.95	0.89	0.91	0.94	0.97
	4	1.01	0.9	0.9	0.91	0.95	0.95	0.85	0.86	0.87	0.92
1000	2	0.97	0.9	0.92	0.94	0.97	0.95	0.89	0.91	0.94	0.97
	4	0.98	0.89	0.87	0.9	0.94	0.96	0.87	0.86	0.89	0.93

m =25

T	r	tr MSE(VEC)					tr MSE(MTV)				
		h					h				
		1	5	10	15	30	1	5	10	15	30
100	10	1.56	1.44	1.34	1.27	1.17	1.19	1.08	1.03	1.02	1.01
	20	1.46	1.28	1.19	1.15	1.1	1.12	1	0.98	0.98	0.99
500	10	1.11	1.06	1.02	1.01	1.01	0.99	0.95	0.96	0.97	0.99
	20	1.11	1.02	0.97	0.97	0.98	0.96	0.84	0.83	0.86	0.91
1000	10	1.03	0.97	0.97	0.98	0.99	0.96	0.91	0.93	0.95	0.98
	20	1.02	0.88	0.86	0.88	0.92	0.94	0.82	0.82	0.85	0.9

Note: See Table 1 for an explanation of the notation.

Table 3: Performance of the VEC and the MTV Approach - Case 3

m =5

T	r	tr MSE(VEC)					tr MSE(MTV)				
		h					h				
		1	5	10	15	30	1	5	10	15	30
100	2	1.56	1.03	1.02	1.03	1.02	1.02	1.01	1	1.01	1.01
	4	1.49	1.01	1.01	1.02	1.03	1	0.98	0.98	0.99	0.99
500	2	1.25	0.98	0.97	0.99	1.01	1	0.96	0.97	0.98	1
	4	1.28	0.94	0.91	0.93	0.95	0.98	0.89	0.9	0.92	0.95
1000	2	1.12	0.95	0.96	0.98	1	1	0.94	0.95	0.97	0.99
	4	1.14	0.92	0.92	0.92	0.96	0.98	0.9	0.9	0.91	0.95

m =25

T	r	tr MSE(VEC)					tr MSE(MTV)				
		h					h				
		1	5	10	15	30	1	5	10	15	30
100	10	1.97	1.48	1.4	1.33	1.21	1.16	1.08	1.04	1.03	1.02
	20	2.04	1.4	1.3	1.25	1.18	1.11	1.03	1	1	1
500	10	1.35	1.06	1.02	1.02	1.01	1.03	1.02	1.03	1.03	1.02
	20	1.36	1.01	0.92	0.93	0.96	0.98	0.9	0.9	0.92	0.95
1000	10	1.2	1	1.03	1.06	1.08	1.01	0.99	1	1.01	1.02
	20	1.19	0.92	0.91	0.93	0.97	0.96	0.86	0.86	0.89	0.94

Note: See Table 1 for an explanation of the notation.

that in small samples, the MTV approach outperforms the VEC approach because it economizes on the number of parameters to be estimated. In Cases 2 and 3, where the error term of the DGP is not identically independently distributed, the MTV, as expected, clearly dominates over the VEC. This result shows that the MTV approach is robust.

(iii) When $r/m = 0.8$, the MTV forecast gain in accuracy relative to the univariate ARIMA forecast as m becomes large. This result indicates that, as already argued by Peña and Poncela (2004), m matters in the relative accuracy of the forecasts.

(iv) The accuracy of the MTV forecast relative to the univariate ARIMA forecast increases when r/m is close to unity.

(v) When h is large, the univariate ARIMA forecast performs as well as the VEC and MTV forecasts. These experimental results are consistent with analytical result (6) and Proposition 2(iii). Our experimental results thus confirm that imposing cointegration does not matter in long-term forecasts but does matter in short- to medium-term forecasts.

In conclusion, based on our experiment, the MTV approach can be recommended as a practical method for forecasting large cointegrated processes when T and/or r/m are large.

6 EMPIRICAL ILLUSTRATIONS

In order to provide an empirical illustration of the MTV approach, we apply it here to forecast the stock prices of pharmaceutical companies listed on the First Section of the Tokyo Stock Exchange. We compare the accuracy of the MTV forecast with that of the VEC forecast and the univariate ARIMA forecast in terms of the trace MSE ratio. The price data are Friday closing prices for the period from January 1990 to December 2000 for a total of 538 weeks. The data came from the Kabuka CD-ROM by Toyo Keizai Shimpo-sha. The first 508 observations were used for estimating the model and the remaining 30 observations for evaluating the accuracy of the forecasts.

In the previous section, we highlighted the interesting finding that when r becomes large while m is fixed, the accuracy of the MTV forecast relative to the univariate ARIMA forecast improves. In order to corroborate this finding, we considered two different sets of 25 arbitrarily chosen pharmaceutical companies with different r.

We first applied the PP test to determine the degree of integration of the 50 series. We could not reject the unit root hypothesis for any of the series and therefore concluded that each series contains a unit root. Next, we sequentially applied the PP test with a 1% significance levels to the principal components. This is the testing procedure for the cointegration rank described in Section 3. The cointegration rank of the first set (denoted as Case A) was found to be 8, whereas that of the second set (denoted as Case B) was 12. In order to compute the VEC forecast, we considered the following VEC model:

$$\Delta y_t = \alpha\beta^{+\prime} y_{t-1}^{+} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \tag{25}$$

where $y_{t-1}^{+} = [1, y_{t-1}']'$, $\beta^{+} = [\rho_0, \beta']'$ and ρ_0 is an $r \times 1$ parameter vector. The SBIC selected $p = 1$, and the maximum likelihood methods developed by Johansen (1998, 1995) were used to estimate coefficient parameters. Moreover, the likelihood ratio test suggested by Johansen (1995) was used to specify the deterministic term. As a result, our model (25) corresponds to (24) which is the DGP for Case 3 in our Monte Carlo experiments.

Case A: Cointegration Rank = 8

Figure 1(a) shows the 25 series of Case A. Figure 2(a) provides the tr MSE(VEC) and the tr MSE(MTV) for the forecasts of 30 periods ahead. The VEC forecast performs poorly relative to the univariate ARIMA forecast, with tr MSE(VEC) ranging from 1.40 to 2.46. On the other hand, the MTV approach yields a better performance than the univariate ARIMAs when $h \geq 4$. Explicitly, tr MSE(MTV) is less than 0.9 for $h = 12$ and 13, indicating clear improvements over the univariate forecasts. To give a typical example, Figure 3(a) provides the forecasts and the actual stock price for one of the 25 pharmaceutical companies, Chugai-Seiyaku. The figure shows that the MTV forecast

Figure 1(a): Stock Prices; Case A

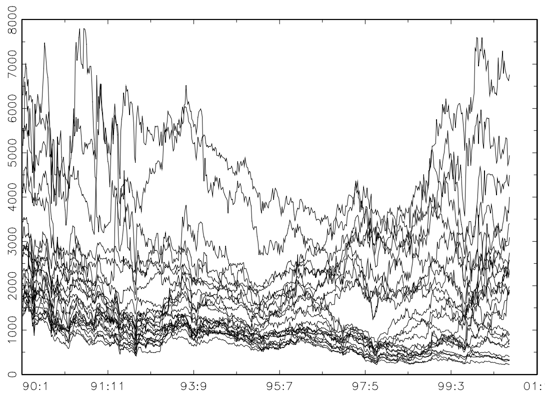
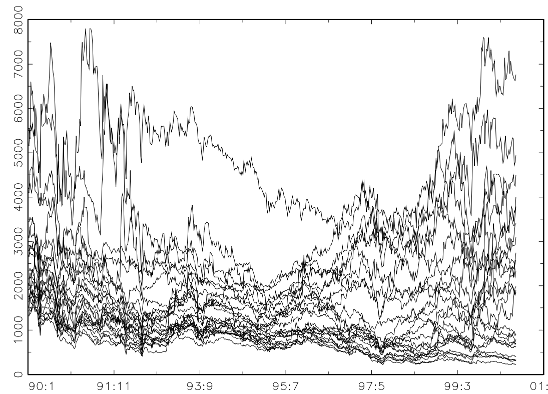


Figure 1(b): Stock Prices; Case B



Note: These figures depict the actual stock prices of the 25 pharmaceutical companies.

works slightly better than the univariate ARIMA forecast. The figure also shows that the VEC forecast performs worse than the univariate ARIMA forecast.

Case B: Cointegration Rank = 12

Figure 1(b) depicts the stock price movements of the other group of 25 companies and Figure 2(b) shows the trace MSE ratios. As in Case A, the VEC forecast performs worse than the MTV forecast, with the values of tr MSE(VEC) generally exceeding 1.5. In contrast, tr MSE(MTV) falls below 0.80 for $h \geq 4$ and reaches 0.66 at its minimum, which means the relative advantage of the MTV forecast over the univariate ARIMA forecast becomes more evident in Case B than in Case A. The MTV forecast works substantially better than the VEC forecast and the univariate ARIMA forecast in this case. This result is consistent with the experimental result that a larger cointegration rank leads to a better performance of the MTV forecasts. Figure 3(b) shows the actual data and the forecasts for Chugai-Seiyaku (Chugai-Seiyaku is included in both data sets). This figure illustrates that the MTV approach works well.

These empirical illustrations indicate that in practice, the MTV approach works well, particularly when the cointegration rank is relatively large in comparison to the number of variables in the process. This result corroborates the experimental results obtained in

Figure 2(a): Trace MSE ratios; Case A

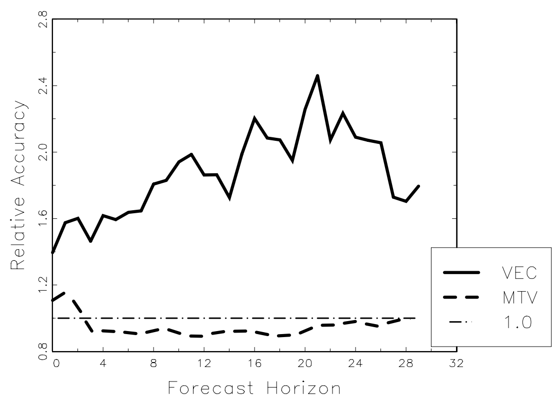
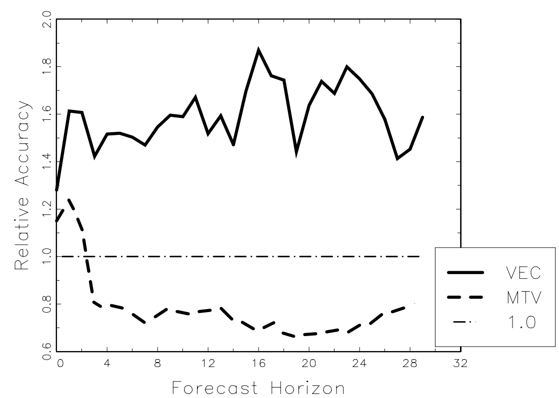


Figure 2(b): Trace MSE ratios; Case B



Note: These figures depict the trace MSE ratios. "VEC" shows the trace MSE ratio of the VEC forecast to the univariate ARIMA forecast. "MTV" shows the trace MSE ratio of the MTV forecast to the univariate ARIMA forecast.

Figure 3(a): Actual and Forecasts; Case A

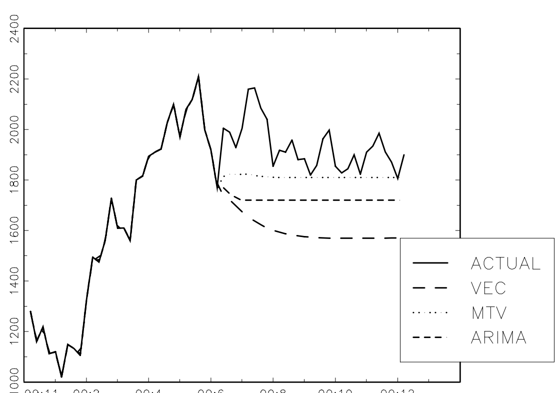
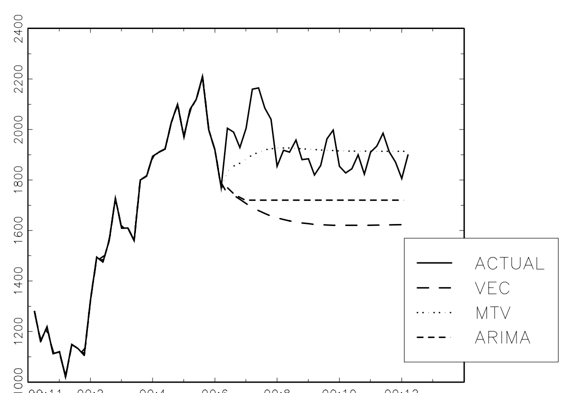


Figure 3(b): Actual and Forecasts; Case B



Note: Note: These figures depict the actual stock price and forecasts for a selected company, Chugai-Seiyaku. "ACTUAL" represents the actual data, while "VEC," "MTV" and "ARIMA" represent the VEC forecast, the MTV forecast, and the univariate ARIMA forecast, respectively.

Section 5. The poor performance of the VEC approach may be attributable to the fact that there is no guarantee that the data can be represented by a VAR model with iid normal disturbances. Further, the model for the VEC forecast contained a deterministic term and the error terms may be severely heteroscedastic, as implied by Figures 1(a) and 1(b). As shown in Section 5, a deterministic term and heteroscedasticity negatively affect the VEC forecast.

In order to check the robustness of these results, we also applied the VEC and the MTV approach to the stock prices of 20 companies in the steel industry. The results were very similar to our exercise with the pharmaceutical companies. That is, the MTV forecast almost always outperforms the VEC and the univariate ARIMA forecast, and the advantage of the MTV forecast increases with r. In addition, the VEC forecast is outperformed by the univariate ARIMA forecast, even when r is sufficiently large. Due to space limitations, we omit the details.

7 CONCLUDING REMARKS

In this paper, we proposed an approach, which we call the MTV approach, for forecasting large cointegrated processes. The MTV approach is free from two problems typically associated with forecasting large cointegrated processes: the difficulty of finding the cointegration rank, and the small sample size relative to the large number of parameters to be estimated. An additional merit of our approach is that it is not necessary, as in a VEC model, to approximate the process or to assume iid normal disturbances.

The results of our Monte Carlo experiments indicated that the MTV forecast works well in comparison to the VEC forecast and the univariate ARIMA forecast. That is, imposing cointegration and economizing the number of parameters to be estimated help to improve the accuracy of the MTV forecast. When T and/or the ratio r/m are large, the MTV forecast becomes more accurate than the other forecasts. In addition, if r/m is large, the MTV forecast gains accuracy as m becomes large. In our empirical illustration, the MTV approach performed rather well: the MTV forecast outperformed the VEC forecast and the univariate ARIMA forecast, thus corroborating the results of our Monte

Carlo experiments.

The MTV approach also performed well in our empirical illustrations and, in fact, the MTV forecast outperformed the VEC forecast and the univariate ARIMA forecast, corroborating the results of our Monte Carlo experiments.

APPENDIX

Proof of Proposition 2

We first prove Proposition 2(i). For large T, multiplying (15) by β' gives

$$\begin{aligned}\beta'\hat{\hat{y}}_{t+h} &= \beta'\beta_{\perp}Q_{\perp}\widehat{B'_{(m-r)}\bar{y}}_{t+h} + \beta'\beta Q\widehat{B'_{(r)}\bar{y}}_{t+h} \\ &= 0 + \beta'\beta Q\widehat{B'_{(r)}\bar{y}}_{t+h}.\end{aligned}$$

Since $\widehat{B'_{(r)}\bar{y}}_{t+h}$ are forecasts from stationary ARMA models, we have the long-term forecast:

$$\lim_{h\to\infty}\beta'\hat{\hat{y}}_{t+h} = 0. \tag{26}$$

This completes the proof of Proposition 2(i).

In order to prove Proposition 2(ii), we introduce the following proposition:

Proposition 3 (Chigira and Yamamoto (2006)):

$$\lim_{h\to\infty}\frac{\text{trace}(E(\ddot{e}_{t+h}\ddot{e}'_{t+h}))}{\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))} = 1, \tag{27}$$

where $\ddot{e}_{t+h} = y_{t+h} - \ddot{y}_{t+h}$ *and* $\ddot{y}_{t+h}$ *is a forecast which can be expressed as*

$$\ddot{y}_{t+h} = (t+h)\mu + v_{t+h}, \tag{28}$$

where v_{t+h} *is an* $m \times 1$ *random vector, which possibly depends on a set of variables observed at date* t*, with* $\text{trace}E(v_{t+h}v'_{t+h}) = O(1)$.

Proof: See Chigira and Yamamoto (2006).

Proposition 3 extends Christoffersen and Diebold's (1998) result, i.e. (6), and shows that under the trace MSE ratio, a forecast given by (28) and the cointegrated system forecast are equally poor in the long-run. We find that many forecasts, including the univariate ARIMA forecast, can be written in the form of (28). Further, a forecast $\hat{\hat{y}}_{t+h} = (t+h)\mu$, which imposes neither integration nor cointegration, is given by (28) with $v_{t+h} = 0$, which indicates that (28) forms a wide class of forecasts. We find that if T is sufficiently large, the MTV forecast also can be expressed as (28) because we can construct a consistent estimator for a constant drift, and the stochastic part of the MTV forecast, i.e. $\beta_\perp Q_\perp \widehat{B'_{(m-r)}\bar{y}}_{t+h} + \beta Q \widehat{B'_{(r)}\bar{y}}_{t+h}$, does not diverge as h becomes large. Thus, from Proposition 3, we obtain the desired result:

$$\lim_{h\to\infty} \frac{\text{trace}(E(\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}))}{\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))} = 1. \tag{29}$$

We now provide the proof of Proposition 2(iii). Use (29) and (6) to obtain

$$\begin{aligned}\lim_{h\to\infty} \frac{\text{trace}(E(\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))} &= \lim_{h\to\infty} \frac{\text{trace}(E(\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}))/\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))}{\text{trace}(E(\tilde{e}_{t+h}\tilde{e}'_{t+h}))/\text{trace}(E(\hat{e}_{t+h}\hat{e}'_{t+h}))} \\ &= \frac{1}{1}.\end{aligned}$$

This completes the proof of Proposition 2(iii).

To prove Proposition 2(iv), we decompose $\hat{\hat{e}}_{t+h}$ as

$$\hat{\hat{e}}_{t+h} = \hat{e}_{t+h} + (\hat{y}_{t+h} - \hat{\hat{y}}_{t+h}). \tag{30}$$

Because ε_t is an iid process and $\hat{y}_{t+h}$ and $\hat{\hat{y}}_{t+h}$ only depend on current and past ε_t, $\hat{e}_{t+h}$ is orthogonal to the terms in the parentheses in (30). Thus, we obtain

$$\begin{aligned}\text{trace}(E(\beta'\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}\beta)) =& \text{trace}(E(\beta'\hat{e}_{t+h}\hat{e}'_{t+h}\beta)) \\ &+ \text{trace}(E[\beta'(\hat{y}_{t+h} - \hat{\hat{y}}_{t+h})(\hat{y}_{t+h} - \hat{\hat{y}}_{t+h})'\beta]).\end{aligned} \tag{31}$$

We first examine the property of the second term in (31) for large h. Use (5) and Proposition 2(i) to obtain

$$\beta'(\hat{y}_{t+h} - \hat{\hat{y}}_{t+h}) \approx 0,$$

where the approximation holds when h is large. Therefore, the second term of (31) disappears as h becomes large, and we obtain

$$\lim_{h\to\infty} \text{trace}(E(\beta'\hat{e}_{t+h}\hat{e}'_{t+h}\beta)) = \lim_{h\to\infty} \text{trace}(E(\beta'\hat{\hat{e}}_{t+h}\hat{\hat{e}}'_{t+h}\beta)). \tag{32}$$

As shown in Engle and Yoo (1987), $\text{trace}(E(\beta'\hat{e}_{t+h}\hat{e}'_{t+h}\beta))$ is finite even if h is large. On the other hand, Christoffersen and Diebold (1998) showed that

$$\lim_{h\to\infty} \text{trace}(E(\beta'\tilde{e}_{t+h}\tilde{e}'_{t+h}\beta)) = \lim_{h\to\infty} \text{trace}(E(\beta'\hat{e}_{t+h}\hat{e}'_{t+h}\beta)) + \text{trace}(\beta' F\beta) < \infty, \tag{33}$$

where F is a positive definite matrix. Then from (32) and (33), we obtain the desired result.

References

Bai, J. and S. Ng (2004): "A PANIC attack on unit roots and cointegration," *Econometrica*, 72, 1127-1177.

Chigira, H. (2006): "A test for cointegration rank based on principal component analysis,"*Applied Economics Letters*, forthcoming.

Chigira, H. and T. Yamamoto (2006): "Cointegration, integration, and long-term forecasting," Hi-Stat Discussion Paper Series No.148, Institute of Economic Research, Hitotsubashi University.

Christoffersen, P.F. and F.X. Diebold (1997): "Cointegration and long-horizon forecasting," Technical Working Paper No. 217, National Bureau of Economic Research, Cambridge, MA.

Christoffersen, P.F. and F.X. Diebold (1998): "Cointegration and long-horizon forecasting," *Journal of Business and Economic Statistics*, 16, 450-458.

Clements, M.P. and D.F. Hendry (1993): "On the limitations of comparing mean square forecast errors," *Journal of Forecasting*, 12, 617-637.

Clements, M.P. and D.F. Hendry (1995): "Forecasting in cointegrated systems," *Journal of Applied Econometrics*, 10, 127-146.

Escribano, A. and D. Peña (1994): "Cointegration and common factors," *Journal of Time Series Analysis*, 15, 577-586.

Engle, R.F. and S. Yoo (1987): "Forecasting and testing in cointegrated systems," *Journal of Econometrics*, 35, 143-159.

Harris, D. (1997): "Principal components analysis of cointegrated time series," *Econometric Theory*, 13, 529-557.

Johansen, S. (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica*, 59, 1551-1580.

Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Kariya, T. (1988): "MTV model and its application to the prediction of stock prices," *Proceedings of the Second International Tampere Conference in Statistics*, 161-176.

Lin, J.L. and R. Tsay (1996): "Co-integration constraint and forecasting: An empirical examination," *Journal of Applied Econometrics*, 114, 519-538.

Osterwald-Lenum, M. (1992): "A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics," *Oxford Bulletin of Economics and Statistics*, 54, 461-472.

Park, J.Y. and P.C.B. Phillips (1988): "Statistical inference in regressions with integrated processes, part 1," *Econometric Theory*, 4, 468-498.

Peña, D. and P. Poncela (2004): "Forecasting with nonstationary dynamic factor models," *Journal of Econometrics*, 119, 291-321.

Phillips, P.C.B. (1987): "Time series regression with a unit root," *Econometrica*, 55, 277-301.

Phillips, P.C.B. and P. Perron (1988): "Testing for a unit root in time series regression," *Biometrika*, 75, 335-346.

Reinsel, G.C. and S.K. Ahn (1992) "Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting," *Journal of Time Series Analysis*, 75, 335-383.

Said, S.E., and D.A. Dickey (1984): "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, 71, 599-607.