THE EFFECT OF ESTIMATING PARAMETERS ON LONG-TERM FORECASTS FOR COINTEGRATED SYSTEMS

Hiroaki CHIGIRA — Department of Economics, Tohoku University

Taku YAMAMOTO — College of Economics, Nihon University

August 2010

1 INTRODUCTION

• **Forecasting in Cointegrated Processes in the Long-run**

(1) **System forecast (VEC (Vector Error Corrction) forecast) "with" cointegration constraint.**

(2) **Other forecasts: Univariate ARIMA forecast, Unrestricted VAR forecast and so on "without" cointegration and/or integration constraints.**

System forecast is optimal in the short-run.

What happens in the long-run?
Does cointegration or integration constraint matter in the long-run forecast?

⇒ Engle and Yoo (1987): Cointegration matters in the long-run forecast.
Lin and Tsay (1996): Proper order of cointegration matters.
(Experimental Results.)

⇒ Cristoffersen and Diebold (1998):
Only integration, bu not cointegration, matters in the long-run.
(Analytical Result.)

• Conclusions of the present paper:

When the parameters are known

Neither conintegration nor integration matters in the long-run forecast of cointegrated processes, based upon the MSE ratio criteria.

When the parameters are estimated

The same conclusion applies with a different reason.

The estimation of the drift term is crucial.

The univariate ARIMA forecast performs better than the cointegrated system forecast. The drift is estimated by time average of differenced data

• Organization of the Paper

Sec.2: Forecasts with Known Parameters.

Sec.3: Forecasts with Estimated Parameters.

Sec.4: Monte Carlo Experimente.

Sec.5: Concluding Remarks.

2 FORECASTS WITH KNOWN PARAMETERS

2.1 Model and System Forecast

- Model: m-variate VMA(∞) Model

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

where L is the lag operator,
$y_t = [y_{1t}.y_{2t}, .., y_{mt}]'$,
μ is a vector of constants,
C_i is a $m \times m$ matrix with absolute summability of $\{sC_s\}_{s=0}^{\infty}$,
ε_t is a $(m \times 1)$ iid$(0, \Sigma)$ process with finite fourth moments.

- Cointegration Rank is r
  β is $m \times r$ cointegration matrix.
  $z_t = \beta' y_t$ is zero-mean $I(0)$ process, or
  $\beta' C(1) = 0$.

- h-period Ahead Forecast

y_{T+h} is expressed as

$$y_{T+h} = h\mu + \sum_{t=1}^{h} \sum_{j=0}^{h-t} C_j \varepsilon_{T+t} + y_T + \sum_{t=1}^{T} \sum_{j=1}^{h} C_{T-t+j} \varepsilon_t \tag{2}$$

The second term in the right-hand-side in the above expressions represents the future errors which are unknown at time T.

The (cointegrated) system forecast or the optimal forecast is given from (2)

$$\hat{y}_{T+h} = h\mu + y_T + \sum_{t=1}^{T} \sum_{j=1}^{h} C_{T-t+j} \varepsilon_t \tag{3}$$

The forecast error is obtained:

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

They show that the trace MSE of $\hat{e}_{T+h}$ diverges as h goes to ∞:

$$\text{trace MSE}(\hat{e}_{T+h}) \equiv \text{trace}(E(\hat{e}_{T+h}\hat{e}'_{T+h})) = O(h). \tag{5}$$

- Accuracy Measure

Accuracy measure: trace MSE

$$\text{trace MSE} = \text{trace}(\boldsymbol{E}(e_{T+h}^{(j)} e_{T+h}^{(j)\prime}))$$

where $e_{T+h}^{(j)}$ is the forecast error of the h period ahead of the j-th method.

Our Criteria for Comparison:

Relative accuracy of forecasts: trace MSE ratio

$$\frac{\text{trace MSE}(e_{T+h}^{(2)})}{\text{trace MSE}(e_{T+h}^{(1)})},$$

where $e_{T+h}^{(j)}$ is the forecast error of the j-th method.

2.2 Comparison with Other Forecasting Methods

An alternative forecating scheme y^+_{T+h} is expressed as

$$y^+_{T+h} = \hat{y}_{T+h} + (y^+_{T+h} - \hat{y}_{T+h}), \tag{6}$$

where $(y^+_{T+h} - \hat{y}_{T+h})$ is the difference between an altenative scheme y^+_{T+h} and the optimal forecast $\hat{y}_{T+h}$.

Its forecast error is given by

$$e^+_{T+h} = y_{T+h} - y^+_{T+h} = y_{T+h} - \{\hat{y}_{T+h} + (y^+_{T+h} - \hat{y}_{T+h})\} = \hat{e}_{T+h} - (y^+_{T+h} - \hat{y}_{T+h}). \tag{7}$$

In general, we have

$$\text{trace MSE}(y^+_{T+h} - \hat{y}_{T+h}) = O(1) = o(h).$$

Thus, we have

$$\lim_{h\to\infty} \frac{\text{trace MSE}(e^+_{T+h})}{\text{trace MSE}(\hat{e}_{T+h})} = 1. \tag{8}$$

For example, y^+_{T+h} includes

(a) Engle and Yoo's (1987) forecast from an unrestricted VAR (UVAR),
(b) Cristoffersen and Diebold's (1998) forecast from a univariate ARIMA model, and
(c) $y^+_{T+h} = y_T$.

Obviously, y_T satisfies neither cointegration nor integration.

Neither cointegration nor integration matters in the long-run.

3 FORECASTS WITH ESTIMATED PARAMETERS

• Denote some estimators of μ and C_i $(i = 0, 1, ...)$ as μ^* and C_i^* $(i = 0, 1, ...)$, respectively.

• Suppose that μ^* is a consistent estimator of μ and Var $(\mu^*) = O(1/T)$.

• Estimators C_i^* (i=0, 1, ...) are not necessarily consistent estimators of the corresponding C_i (i=0, 1, ...).

Thus, we may note that $C_i^* - C_i$ can be $O(1)$ for any i.

We now denote forecasts with the estimated parameters with superscript *.

For example, we denote the cointegrated system forecast with the estimated parameters μ^* and C_i^* (i=0, 1, ...) as y_{T+h}^* is given as follows:

$$y_{T+h}^* = h\mu^* + y_T + \sum_{t=1}^{T}\sum_{j=1}^{h} C_{T-i-j}^* \varepsilon_i$$

Its forecast error is given as

$$e^*_{T+h} = y_{T+h} - y^*_{T+h} = (\mu - \mu^*)h + \sum_{t=1}^{h}\sum_{j=0}^{h-t} C_j \varepsilon_{T+t} + \sum_{t=1}^{T}\sum_{j=1}^{h}(C_{T-t+j} - C^*_{T-t+j})\varepsilon_t$$
$$= P + Q + R. \quad \text{(say)}$$

We find taht $E[QP'] = E[QR'] = 0$, since Q consists solely of future errors and is uncorrelated with P and R. We further note that, for given T,

$$\begin{aligned} E[PP'] &= \text{MSE}[\mu^*]h^2 = O(h^2), \\ E[QQ'] &= E[RR'] = O(h), \text{ and} \\ E[PR'] &= O(h). \end{aligned}$$

The first term in the right-hand-side $P = (\mu - \mu^*)h$, the error associated with estimation of μ, is dominant for given T.

As h goes to infinity, the trace MSE of e^*_{T+h} is now given by

$$\text{trace MSE}(e^*_{T+h}) \equiv \text{trace}(E(e^*_{T+h}e^{*\prime}_{T+h})) = \text{MSE}[\mu^*]h^2.$$

Proposition: *Suppose that $e^{(1)*}_{T+h}$ and $e^{(2)*}_{T+h}$ are forecast errors of two different estimation methods. Let $\mu^{(1)*}$ and $\mu^{(2)*}$ be their estimators of μ, respectively. Then, we have, for given T,*

$$\lim_{h\to\infty} \frac{\text{trace MSE}(e^{(2)*}_{T+h})}{\text{trace MSE}(e^{(1)*}_{T+h})} = \frac{\text{trace MSE}[\mu^{(2)*}]}{\text{trace MSE}[\mu^{(1)*}]}.$$

Remarks:

- The relative accuracy of long-term forecasts with estimated parameters depends upon the relative MSEs of drift estimators.

- It is interesting to note that the ratio never explodes nor conveges to zero.

- It should be noted that the above result is valid as long as the drift is estimated, even when the true μ is zero.

- It is important to realize that the estimation error of the drift term is crucial for long-term forecasts.

⇒ Neither cointegration nor integration matters in the long-run, when the paremeters of the model are estimated.

4 MONTE CARLO EXPERIMENTS: How Forecasts Perform When Prediction Horizon Increases

4.1 The Monte Carlo Design

As a DGP, consider a simple bivariate cointegrated process as follows:

$$\Delta y_t = \alpha\beta' y_{t-1} + m + \varepsilon_t,$$

where α, β and m are specifically given as

$$\alpha = \begin{bmatrix} -0.4 \\ 0.1 \end{bmatrix},\ \beta = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix},\ m = \begin{bmatrix} 0.24 \\ 0.24 \end{bmatrix},$$

and $\varepsilon_t \sim i.i.d.N(0, I_2)$. Note that the constant (drift) term m is added here.

It can be expressed as a bivariate vector autoregressive (VAR) model:

$$y_t = A y_{t-1} + m + \varepsilon_t,$$

where $A = I_2 + \alpha\beta'$.

We may note that Δy_t can be also be expressed as MA(∞) process.

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

where $C(L) = \sum_{i=0}^{\infty} C_i L^i$, $C_0 = I_2$, $C_i = A^i - A^{i-1}$ $(i = 1, 2, \cdots)$, $\mu = [\mu_1, \mu_2]' = Cm$, and $C = C(1) = \beta_\perp(\alpha'_\perp \beta_\perp)^{-1}\alpha'_\perp$. (See Johansen (1995) for derivation of the moving average representation of Δy_t above.)

In the numerical example, we have

$$\mu = [\mu_1,\ \mu_2]' = [0.4,\ 0.2]'.$$

Throughout our experiment, we set the sample size to 100 and 200, and the number of replications is 4000.

4.2 Experimental Results (Model is correctly specified)

We denote the estimated parameters wirh superscript *.

(i*) The cointegrated system forecast (SYS*), $\hat{y}^*_{T+h}$:

The cointegrated system forecast of y^*_{t+h} with estimated parameters is given by

$$\hat{y}^*_{T+h} = \hat{A}^h y_T + \hat{\mu}(h)\,,$$

where where $\hat{\mu}(h) = (I_2 + \hat{A} + \hat{A}^2 + \cdots + \hat{A}^{(h-1)})\hat{m}$.

Here, the parameters of the VEC model, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{m}$ are estimates by Johansen's maximum likelihood procedure, and we have $\hat{A} = I_2 + \hat{\alpha}\hat{\beta}'$.

Here, we assume that the cointegration rank and the lag length of the model are a priori known.

In this case we have $\hat{\mu} = \hat{C}\hat{m}$ where $\hat{C} = \hat{\beta}_\perp(\hat{\alpha}'_\perp\hat{\beta}_\perp)^{-1}\hat{\alpha}'_\perp$.

(ii*) [The approximate cointegrated system forecast (APPROX*), $\ddot{y}^*_{T+h}$:

The approximate cointegrated system forecast with estimated parameters is given by

$$\ddot{y}^*_{T+h} = \hat{A}^h y_T + h\check{\mu}\,,$$

where, since $E(\Delta y_t = \mu)$ (See Johansen (1995)), $\check{\mu}$ is directly obtained by

$$\check{\mu} = \frac{1}{T-1}\sum_{t=2}^{T} \Delta y_t.$$

(iii*) The univariate ARIMA forecast (ARIMA*), $\tilde{y}^*_{t+h}$:

For an individual series, we can make forecasts from ARIMA$(r, 1, s)$ models, denoted as $\tilde{y}^*_{t+h}$.

We fit an ARMA(r, s) model for an individual series of $\Delta y_t - \breve{\mu} \quad (t = 2, 3, \cdots, T)$, where $\breve{\mu}$ is defined above.

The orders of r and s are determined by SBIC for each replication.

(iv*) The misspecified forecast (MISS*), $\hat{\hat{y}}^*_{T+h}$:

This is the forecast from simple trend models, denoted as $\hat{\hat{y}}^*_{T+h}$:

$$\hat{\hat{y}}^*_{1,T+h} = \hat{\hat{\phi}}_1 + \hat{\hat{\mu}}_1(T + h), \quad \text{and}$$
$$\hat{\hat{y}}^*_{2,T+h} = \hat{\hat{\phi}}_2 + \hat{\hat{\mu}}_2(T + h),$$

where $\hat{\hat{\phi}} = [\hat{\hat{\phi}}_1,\ \hat{\hat{\phi}}_2]'$ and $\hat{\hat{\mu}} = [\hat{\hat{\mu}}_1,\ \hat{\hat{\mu}}_2]'$ are estimated by the OLS estimation of the following models;

$$y_{1,t} = \phi_1 + \mu_1 t + \text{error}, \quad \text{and}$$
$$y_{2,t} = \phi_2 + \mu_2 t + \text{error}.$$

(v) The idle forecast (IDLE), $\bar{y}_{t+h}$:

The idle forecast is defined as, regardless of the forecast horizon h,

$$\bar{y}_{t+h} = [120,\ 60]'.$$

We consider the following four trace MSE ratios:

$$\text{ratio}(\text{APPROX}^*) = \frac{\text{trace MSE}(\ddot{e}^*_{T+h})}{\text{trace MSE}(\hat{e}^*_{T+h})},$$

$$\text{ratio}(\text{ARIMA}^*) = \frac{\text{trace MSE}(\tilde{e}^*_{T+h})}{\text{trace MSE}(\hat{e}^*_{T+h})},$$

$$\text{ratio}(\text{MISS}^*) = \frac{\text{trace MSE}(\hat{\hat{e}}^*_{T+h})}{\text{trace MSE}(\hat{e}^*_{T+h})}, \quad \text{and}$$

$$\text{ratio}(\text{IDLE}^*) = \frac{\text{trace MSE}(\bar{e}_{T+h})}{\text{trace MSE}(\hat{e}^*_{T+h})},$$

where $\hat{e}^*_{T+h}$, $\ddot{e}^*_{T+h}$, $\tilde{e}^*_{T+h}$, $\hat{\hat{e}}^*_{T+h}$ and $\bar{e}_{T+h}$ are the forecast errors of forecasts (i*), (ii*), (iii*) , (iv*), and (v), respectively.

Experimental Results when $T = 100$

Figure 2(a) shows ratio(APPROX*), ratio(ARIMA*) and ratio(MISS*) plotted against forecast horizon h.

Figure 2 Trace MSE Ratios When Paremeters Are Estimated ($\boldsymbol{T} = 100$)

Note:
"ratio(APPROX*)": Trace MSE ratio of APPROX* to SYS*,
"ratio(ARIMA*)" : Trace MSE ratio of ARIMA* to SYS*,
"ratio(MISS*)" : Trace MSE ratio of MISS* to SYS*, and
"ratio(IDLE*)" : MSE ratio of IDLE to SYS*.

In this figure, we can see that they intially larger than unity for small h. That is, the cointegrated system forecast with estimated parameters is best for short-term forecasts.

However, next, they uniformly decrease below unity as h becomes large.

The approximate cointegratesd system forcast (APPROX*), the univariate ARIMA forcast (ARIMA*), and the misspecified forecast (MISS*) are better than the cointegrated system forecast (SYS*) in the long-run.

Actually, we have, for $h = 1000$,

$$\text{ratio(APPROX}^*) = 0.635$$
$$\text{ratio(ARIMA}^*) = 0.581$$
$$\text{ratio(MISS}^*) = 0.702$$

Table 1 shows that diagonal elments of Var[$\breve{\mu}$] in APPROX* and ARIMA* and Var[$\hat{\hat{\mu}}$] in MISS* are smaller than those of Var[$\hat{\mu}$] in SYS*.

Table 1: Mean and Variance of Drift Term μ ($T = 100$)

True μ	mean of $\hat{\mu} = \hat{C}\hat{m}$	(variance of $\hat{\mu} = \hat{C}\hat{m}$)	mean of $\hat{\hat{\mu}}$	(var. ratio of $\hat{\hat{\mu}}$ and $\hat{\mu}$)	mean of $\breve{\mu}$	(var. ratio of $\breve{\mu}$ and $\hat{\mu}$)
0.4	0.391	(0.0461)	0.400	(0.453)	0.398	(0.372)
0.2	0.193	(0.0112)	0.200	(0.590)	0.200	(0.503)

This may happen because we have to estimate many prameters in the case of a VEC model and it gives large variability on the ML estimators. (Note that, asymptotically, $\hat{\mu}$ and $\breve{\mu}$ have the same variance (Johansen,1995, Th.13.7)).

The result conforms with Proposition 2, which states that the trace MSE ratio converges to trace Var ratio of estimators of $\boldsymbol{\mu}$.

While we have been concerened wth long-term forecasts, our result is also infomative on the medium-term forecasts, say, $h = 15 \sim 50$. Actually,

$$\text{ratio(ARIMA}^*\text{)} < 1 \text{ for } h > 35.$$

Figure 2(b) shows ratio(IDLE*) plotted against horizon h.

We have

$$\begin{aligned}
\text{ratio}(\boldsymbol{IDLE}^*) &= \text{trace}\boldsymbol{MSE}(\bar{e}_{T+h})/\text{trace}\boldsymbol{MSE}(\hat{e}^*_{T+h}) \\
&\approx (\mu_1^2 + \mu_2^2)(T+h)^2/\text{trace VAR}[\hat{\boldsymbol{\mu}}]h^2 \\
&\xrightarrow{h\to\infty} (\mu_1^2 + \mu_2^2)/\text{trace VAR}[\hat{\boldsymbol{\mu}}] \\
&= \text{constant.}
\end{aligned}$$

4.3 Case Where A Cointegrated System Model Is Misspecified

We now cosider a VEC model with MA errors:

$$\Delta y_t = \alpha\beta' y_{t-1} + m + \Psi\varepsilon_{t-1} + \varepsilon_t, \text{ and}$$
$$\Psi = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix},$$

where α, β, m and ε_t are the same as in (4.1).

We evaluate five forecasts, SYS*, APPROX*, ARIMA*, MISS*, and IDLE* as in the previous subsection.

Here, we may note that the VEC model is estimated with $p = 1$ as a misspecification.

Figure 4 Trace MSE Ratios When Model Is Misspecified (T=100)

Experimental Results When $T = 100$

Figure 4(a) shows ratio(APPROX*), ratio(ARIMA*) and ratio(MISS*) plotted against forecast horizon h.

We can see that they decrease faster than those in Figure 2(a).

That is, the dominace of the approximate cointegratesd system forecast (APPROX*), the univariate ARIMA forcast (ARIMA*), and the misspecified forecast (MISS*) over the cointegrated system forecast (SYS*) is more apparent in the long-run.

Actually, we now have, for $h = 1000$,

$$\text{ratio(APPROX}^*) = 0.332$$
$$\text{ratio(ARIMA}^*) = 0.259$$
$$\text{ratio(MISS}^*) = 0.314$$

They are much smaller than those in those in the previous subsection.

Table 3 shows means and variances of $\hat{\mu}, \hat{\hat{\mu}}$ and $\breve{\mu}$.

Table 3: Mean and Variance of Drift Term μ When Model Is Misspecified

True μ	mean of $\hat{\mu} = \hat{C}\hat{m}$	(variance of $\hat{\mu} = \hat{C}\hat{m}$)	mean of $\hat{\hat{\mu}}$	(var. ratio of $\hat{\hat{\mu}}$ and $\hat{\mu}$)	mean of $\breve{\mu}$	(var. ratio of $\breve{\mu}$ and $\hat{\mu}$)
0.4	0.424	(4.020)	0.399	(0.117)	0.397	(0.00958)
0.2	0.181	(0.386)	0.199	(0.381)	0.200	(0.0326)

Variances are all larger than those in Table 1. Apparently, $\hat{\mu}$ is most affected by the misspecification compared to other two estimators.

It is interesting to note that for ratio(ARIMA*) we have

$$\begin{aligned}\text{ratio}(\text{ARIMA}^*) &= 0.941 \quad (h = 20),\\ \text{ratio}(\text{ARIMA}^*) &= 0.884 \quad (h = 30), \text{and}\\ \text{ratio}(\text{ARIMA}^*) &= 0.797 \quad (h = 50).\end{aligned}$$

Namely, ARIMA* is noticeably better than SYS* for medium-term forecasts, say, $h = 20 \sim 50$.

Actually, ARIMA* almost unifromly dominates SYS* for all h in this case.

Since in practical situations, the msspecification is likely to occur, we recommend to use ARIMA* because of its simpleness and robustness.

5 CONCLUSION

Known Prameters:

- **Neither cointegration nor integration matters in long-term forecasts.**

Estimated Prameters:

- **Accuracy of the estimation of the drift term is crucial in long-term forecasts. Again, neither cointegration nor integration matters in long-term forecasts.**

- **Namely, the relative accuracy of various long-term forecasts depends upon the relative magnitude of MSEs of estimators of the drift term.**

- **The univariate ARIMA forecast whose drift term is estimated by the simple time average of differenced data, is better than cointegrated sytem forecasts whose parameters are estimated by the conventional Johansen's maximum likelihood.**

- **The dominance of the univariate ARIMA forecast over the cointegrated system forecasts can happen even in medium-term forecasts, say, $h = 20 \sim 50$. Our result is informative on medium-term forecasts.**