Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

# Cointegration and Long-Horizon Forecasting

Peter F. Christoffersen & Francis X. Diebold

**To cite this article:** Peter F. Christoffersen & Francis X. Diebold (1998) Cointegration and Long-Horizon Forecasting, Journal of Business & Economic Statistics, 16:4, 450-456, DOI: 10.1080/07350015.1998.10524784

**To link to this article:** https://doi.org/10.1080/07350015.1998.10524784

# Cointegration and Long-Horizon Forecasting

**Peter F. Christoffersen**
Faculty of Management McGill University, Montreal, Quebec H3A 1G5, Canada
(christop@management.mcgill.ca)

**Francis X. Diebold**
Department of Economics, University of Pennsylvania, Philadelphia, PA 19104-6297, and National Bureau of Economic Research, Cambridge, MA 02138 (fdiebold@mail.sas.upenn.edu)

We consider the forecasting of cointegrated variables, and we show that at long horizons nothing is lost by ignoring cointegration when forecasts are evaluated using standard multivariate forecast accuracy measures. In fact, simple univariate Box–Jenkins forecasts are just as accurate. Our results highlight a potentially important deficiency of standard forecast accuracy measures—they fail to value the maintenance of cointegrating relationships among variables—and we suggest alternatives that explicitly do so.

KEY WORDS: Integration; Loss function; Prediction; Unit root.

Cointegration implies restrictions on the low-frequency dynamic behavior of multivariate time series. Thus, imposition of cointegrating restrictions has immediate implications for the behavior of long-horizon forecasts, and it is widely believed that imposition of cointegrating restrictions, when they are in fact true, will produce superior long-horizon forecasts. This view stems from the theoretical result that long-horizon forecasts from cointegrated systems satisfy the cointegrating relationships exactly and the related result that only the cointegrating combinations of the variables can be forecast with finite long-horizon error variance. Moreover, it appears to be supported by several independent Monte Carlo analyses (e.g., Engle and Yoo 1987; Reinsel and Ahn 1992; Clements and Hendry 1993; Lin and Tsay 1996).

This article grew out of an attempt to reconcile the popular intuition just sketched, which seems sensible, with a competing conjecture, which also seems sensible. Forecast enhancement from exploiting cointegration comes from using information in the current deviations from the cointegrating relationships. That is, knowing whether and by how much the cointegrating relations are violated today is valuable in assessing where the variables will go tomorrow because deviations from cointegrating relations tend to be eliminated. Although the current value of the error-correction term clearly provides information about the likely *near-horizon* evolution of the system, however, it seems unlikely that it provides information about the *long-horizon* evolution of the system because the long-horizon forecast of the error-correction term is always 0. (The error-correction term, by construction, is covariance stationary with a zero mean.) From this perspective, it seems unlikely that cointegration could be exploited to improve long-horizon forecasts.

Motivated by this apparent paradox, we provide a precise characterization of the implications of cointegration for long-horizon forecasting. Our work is closely related to important earlier contributions of Clements and Hendry (1993, 1994, 1995) and Banerjee, Dolado, Galbraith, and Hendry (1993, pp. 278–285), who compared forecasts from a true vector autoregression (VAR) to forecasts from a misspecified VAR in differences, whereas we compare the true forecasts to exact forecasts from correctly specified but univariate representations. We focus explicitly and exclusively on forecasting, and we obtain several new theoretical results that sharpen the interpretation of existing Monte Carlo results. Moreover, our motivation is often very different. Rather than focusing, for example, on loss functions invariant to certain linear transformations of the data, we take the opposite view that loss functions—like preferences—are sovereign and explore in detail how the effects of imposing cointegration on long-horizon forecasts vary fundamentally with the loss function adopted. In short, our results and theirs are highly complementary.

We proceed as follows. In Section 1 we show that, contrary to popular belief, nothing is lost by ignoring cointegration when long-horizon forecasts are evaluated using standard accuracy measures; in fact, even *univariate* Box–Jenkins forecasts are equally accurate. In Section 2, we address a potentially important deficiency of standard forecast accuracy measures highlighted by our analysis—they fail to value the maintenance of cointegrating relationships among variables—and we suggest alternative accuracy measures that explicitly do so. In Section 3, we consider forecasting from models with estimated parameters, and we use our results to clarify the interpretation of several well-known Monte Carlo studies. We conclude in Section 4.

## 1. MULTIVARIATE AND UNIVARIATE FORECASTS OF COINTEGRATED SERIES

In this section we establish notation, recall standard results on multivariate forecasts of cointegrated variables, add new results on univariate forecasts of cointegrated vari-


Journal of Business & Economic Statistics
October 1998, Vol. 16, No. 4

ables, and compare the two. First, let us establish some notation.

Assume that the $N \times 1$ vector process of interest is generated by $(1-L)x_t = \mu + C(L)\varepsilon_t$, where $\mu$ is a constant drift term, $C(L)$ is an $N \times N$ matrix lag operator polynomial of possibly infinite order, and $\varepsilon_t$ is a vector of iid innovations. Then, under regularity conditions, the existence of $r$ linearly independent cointegrating vectors is equivalent to $\text{rank}(C(1)) = N - r$, and the cointegrating vectors are given by the rows of the $r \times N$ matrix $\alpha'$, where $\alpha' C(1) = \alpha'\mu = 0$. That is, $z_t = \alpha' x_t$ is an $r$-dimensional *stationary* zero-mean time series. We will assume that the system is in fact cointegrated, with $0 < \text{rank}(C(1)) < N$. For future reference, note that, following Stock and Watson (1988), we can use the decomposition $C(L) = C(1) + (1-L)C^*(L)$, where $C_j^* = -\sum_{i=j+1}^{\infty} C_i$, to write the system in "common-trends" form, $x_t = \mu t + C(1)\xi_t + C^*(L)\varepsilon_t$, where $\xi_t = \sum_{i=1}^{t} \varepsilon_i$.

We will compare the accuracy of two forecasts of a multivariate cointegrated system that are polar extremes in terms of cointegrating restrictions imposed—first, forecasts from the multivariate model and, second, forecasts from the implied univariate models. Both forecasting models are correctly specified from a *univariate* perspective, but one imposes the cointegrating restrictions and allows for correlated error terms across equations and the other does not.

We will make heavy use of a ubiquitous forecast accuracy measure, mean squared error (MSE), the multivariate version of which is $\text{MSE} = E(e'_{t+h} K e_{t+h})$, where $K$ is an $N \times N$ positive definite symmetric matrix and $e_{t+h}$ is the vector of $h$-step-ahead forecast errors. MSE of course depends on the weighting matrix $K$. It is standard to set $K = I$, in which case $\text{MSE} = E(e'_{t+h} e_{t+h}) = \text{trace}(\Sigma_h)$, where $\Sigma_h = \text{var}(e_{t+h})$. We call this the "trace MSE" accuracy measure. To compare the accuracy of two forecasts, say 1 to 2, it is standard to examine the ratio $\text{trace}(\Sigma_h^1)/\text{trace}(\Sigma_h^2)$, which we call the "trace MSE ratio."

## 1.1 Forecasts From the Multivariate Cointegrated System

Now we review standard results (required by our subsequent analysis) on multivariate forecasting in cointegrated systems. For expanded treatments, see Engle and Yoo (1987) and Lin and Tsay (1996).

From the moving average representation, we can unravel the vector process recursively from time $t+h$ to time 1 and write

$$x_{t+h} = (t+h)\mu + \sum_{i=1}^{t} \sum_{j=0}^{t+h-i} C_j \varepsilon_i + \sum_{i=1}^{h} \sum_{j=0}^{h-i} C_j \varepsilon_{t+i},$$

from which the $h$-step-ahead forecasts are easily calculated as

$$\hat{x}_{t+h} = (t+h)\mu + \sum_{i=1}^{t} \sum_{j=0}^{t+h-i} C_j \varepsilon_i.$$

From the fact that

$$\lim_{h \to \infty} \sum_{j=0}^{t+h-i} C_j = C(1),$$

we get that

$$\lim_{h \to \infty} \alpha' \hat{x}_{t+h} = 0$$

so that the cointegrating relationship is satisfied exactly by the long-horizon system forecasts. This is the sense in which long-horizon forecasts from cointegrated systems preserve the long-run multivariate relationships exactly.

We define the $h$-step-ahead forecast error from the multivariate system as $\hat{e}_{t+h} = x_{t+h} - \hat{x}_{t+h}$. The forecast errors from the multivariate system satisfy

$$\hat{e}_{t+h} = \sum_{i=1}^{h} \sum_{j=0}^{h-i} C_j \varepsilon_{t+i},$$

so the variance of the $h$-step-ahead forecast error is

$$\text{var}[\hat{e}_{t+h}] = \sum_{i=1}^{h} \left[ \left( \sum_{j=0}^{h-i} C_j \right) \Omega \left( \sum_{j=0}^{h-i} C'_j \right) \right],$$

where $\Omega$ is the variance of $\varepsilon_t$.

From the definition of $\hat{e}_{t+h}$, we can also see that the system forecast errors satisfy

$$\hat{e}_{t+h} - \hat{e}_{t+h-1} = \sum_{i=1}^{h} C_{h-i} \varepsilon_{t+i} = C(L)\varepsilon_{t+h},$$

where the last equality holds if we take $\varepsilon_j = 0$ for all $j < t$. That is, when we view the system forecast-error process as a function of the forecast horizon, $h$, it has the same stochastic structure as the original process, $x_t$, and therefore is integrated and cointegrated. Consequently, the variance of the $h$-step-ahead forecast errors from the cointegrated system is of order $h$ that is increasing at the rate $h$, $\text{var}[\hat{e}_{t+h}] = O(h)$. In contrast, the cointegrating combinations of the system forecast errors, just as the error-correction process $z_t$, will have finite variance for large $h$,

$$\lim_{h \to \infty} \text{var}[\alpha' \hat{e}_{t+h}] = \alpha' Q \alpha < \infty,$$

where the matrix $Q$ is a constant function of the stationary component of the forecast error. Although individual series can only be forecast with increasingly wide confidence intervals, the cointegrating combination has a confidence interval of finite width, even as the forecast horizon goes to infinity.

## 1.2 Forecasts From the Implied Univariate Representations

Now consider ignoring the multivariate features of the system, forecasting instead using the implied univariate representations. We can use Wold's decomposition theorem and write for any series (the $n$th, say),

$$(1-L)x_{n,t} = \mu_n + \sum_{j=0}^{\infty} \theta_{n,j} u_{n,t-j},$$

where $\theta_{n,0} = 1$ and $u_{n,t}$ is white noise. It follows from this expression that the univariate time $t$ forecast for period $t+h$ is

$$\tilde{x}_{n,t+h} = h\mu_n + x_{n,t} + \left(\sum_{i=1}^{h} \theta_{n,i}\right) u_{n,t} + \left(\sum_{i=2}^{h+1} \theta_{n,i}\right) u_{n,t-1} + \cdots.$$

Using obvious notation, we can write $\tilde{x}_{n,t+h} = h\mu_n + x_{n,t} + \theta_n(L)u_{n,t}$, and stacking the $N$ series we have $\tilde{x}_{t+h} = h\mu + x_t + \Theta(L)u_t$, where $\Theta(L)$ is a diagonal matrix polynomial with the individual $\theta_n(L)$'s on the diagonal.

Now let us consider the errors from the univariate forecasts. We will rely on the following convenient orthogonal decomposition: $\tilde{e}_{t+h} \equiv x_{t+h} - \tilde{x}_{t+h} = (x_{t+h} - \hat{x}_{t+h}) + (\hat{x}_{t+h} - \tilde{x}_{t+h}) = \hat{e}_{t+h} + (\hat{x}_{t+h} - \tilde{x}_{t+h})$. Recall that the system forecast is

$$\hat{x}_{t+h} = \mu(t+h) + \sum_{i=1}^{t} \sum_{j=0}^{t+h-i} C_j \varepsilon_i \approx \mu(t+h) + C(1)\xi_t,$$

where the approximation holds as $h$ gets large. Using univariate forecasts, the decomposition for $\tilde{e}_{t+h}$, and the approximate long-horizon system forecast, we get $\tilde{e}_{t+h} \approx \hat{e}_{t+h} + (\mu(t+h) + C(1)\xi_t) - (x_t + \mu h + \Theta(L)u_t)$. Now insert the common trends representation for $x_t$ to get $\tilde{e}_{t+h} \approx \hat{e}_{t+h} + \mu(t+h) + C(1)\xi_t - (\mu t + C(1)\xi_t + C^*(L)\varepsilon_t + \mu h + \Theta(L)u_t)$, and finally cancel terms to get $\tilde{e}_{t+h} \approx \hat{e}_{t+h} - (C^*(L)\varepsilon_t + \Theta(L)u_t)$.

Notice that the $\varepsilon_t$'s are serially uncorrelated and the $u_t$'s only depend on current and past $\varepsilon_t$'s; thus, $\hat{e}_{t+h}$ is orthogonal to the terms in the parentheses. Notice also that the parenthesis term is just a sum of stationary series and is therefore stationary; furthermore, its variance is constant as the forecast horizon $h$ changes. We can therefore write the long-horizon variance of the univariate forecasts as $\text{var}(\tilde{e}_{t+h}) = \text{var}(\hat{e}_{t+h}) + O(1) = O(h) + O(1) = O(h)$, which is of the same order of magnitude as the variance of the *system* forecast errors. Thus, in the simple MSE sense, the system and univariate forecasts are equally poor: The MSE losses of both sets of forecasts diverge as the horizon increases. Furthermore, because the dominating terms in the numerator and denominator are identical, the trace MSE ratio goes to 1 as formalized in the following proposition:

*Proposition 1.*

$$\lim_{h\to\infty} \frac{\text{trace}(\text{var}(\tilde{e}_{t+h}))}{\text{trace}(\text{var}(\hat{e}_{t+h}))} = 1.$$

When comparing accuracy using the trace MSE ratio, the univariate forecasts perform as well as the cointegrated system forecasts as the horizon gets large. This is the opposite of the folk wisdom—it turns out that imposition of cointegrating restrictions helps at short, but not long, horizons. Quite simply, when accuracy is evaluated with the trace MSE ratio, there is no long-horizon benefit from imposing cointegration; all that matters is getting the level of *integration* right.

Proposition 1 provides the theoretical foundation for the results of Hoffman and Rasche (1996), who found in an extensive empirical application that imposing cointegration does little to enhance long-horizon forecast accuracy, and Brandner and Kunst (1990), who suggested that, when in doubt about how many unit roots to impose in a multivariate long-horizon forecasting model, it is less harmful to impose too many than to impose too few. A similar result can be obtained by taking the ratio of Clements and Hendry's (1995) formulas for the MSE at horizon $h$ from the system forecasts and the MSE of forecasts that they construct that correspond approximately to those from a misspecified VAR in differences.

Now let us consider the variance of cointegrating combinations of univariate forecast errors. Previously we recounted the Engle–Yoo (1987) result that the cointegrating combinations of the system forecast errors have finite variance as the forecast horizon gets large. Now we want to look at the same cointegrating combinations of the *univariate* forecast errors. From our earlier derivations, it follows that $\alpha'\tilde{e}_{t+h} \approx \alpha'\hat{e}_{t+h} - (\alpha'C^*(L)\varepsilon_t + \alpha'\Theta(L)u_t)$. Again we can rely on the orthogonality of $\alpha'\hat{e}_{t+h}$ to the terms in the parentheses. The first term, $\alpha'\hat{e}_{t+h}$, has finite variance, as discussed previously. So too do the terms in the parentheses because they are linear combinations of stationary processes. Thus, we have the following proposition.

*Proposition 2.*

$$\begin{aligned}\text{var}(\alpha'\tilde{e}_{t+h}) &= \alpha'Q\alpha + \alpha'\text{var}(C^*(L)\varepsilon_t + \Theta(L)u_t)\alpha \\ &= O(1).\end{aligned}$$

The cointegrating combinations of the long-horizon errors from the univariate forecasts, which completely ignore cointegration, *also have finite variance.* Thus, it is in fact not imposition of cointegration on the forecasting system that yields the finite variance of the cointegrating combination of the errors; rather it is the cointegration property inherent in the system itself, which is partly inherited by the correctly specified univariate forecasts.

In our study thus far, to guarantee a fair comparison of the forecasting performance of the multivariate (cointegrated system) and univariate forecasts, the univariate forecast model was chosen as the one implied from the underlying true multivariate model, thus inevitably inheriting certain of its features. Because all analysis was in population, estimation error played no part in the results, but in Section 3 we will incorporate the effects of parameter estimation uncertainty. To guarantee a fair comparison there as well, the univariate forecast model *order* [e.g., ARIMA(2, 1, 1)] will continue to be the one implied by the underlying true multivariate model, but the actual parameters will be estimated separately by maximizing the univariate likelihood.

## 2. ACCURACY MEASURES AND COINTEGRATION

### 2.1 Accuracy Measures I: Trace MSE

We have seen that long-horizon univariate forecasts of cointegrated variables (which ignore cointegrating restrictions) are just as accurate as their system counterparts (which explicitly impose cointegrating restrictions) when accuracy is evaluated using the standard trace MSE criterion. So on traditional grounds there is no reason to prefer long-horizon forecasts from the cointegrated system.

One might argue, however, that the system forecasts are nevertheless more appealing because "the forecasts of levels of co-integrated variables will 'hang together' in a way likely to be viewed as sensible by an economist, whereas forecasts produced in some other way, such as by a group of individual, univariate Box–Jenkins models, may well not do so" (Granger and Newbold 1986, p. 226). But as we have seen, univariate Box–Jenkins forecasts *do* hang together if the variables are cointegrated—the cointegrating combinations, and only the cointegrating combinations, of univariate forecast errors have finite variance.

## 2.2 Accuracy Measures II: Trace MSE in Forecasting the Cointegrating Combinations

The long-horizon system forecasts, however, do a better job of satisfying the cointegrating restrictions than do the univariate forecasts—the long-horizon system forecasts *always* satisfy the cointegrating restrictions, whereas the long-horizon univariate forecasts do so only on average. From Proposition 2 it is clear that although the cointegrating combinations of both the univariate and system forecast errors have finite variance, the variance of the cointegrating combination of the univariate errors is larger.

Such effects are lost on standard accuracy measures like trace MSE, however, because the loss functions that underlie them do not explicitly value maintaining the multivariate long-run relationships of long-horizon forecasts. The solution is obvious—if we value maintenance of the cointegrating relationship, then so too should the loss functions underlying our forecast accuracy measures. One approach, in the spirit of Granger (1996), is to focus on forecasting the cointegrating combinations of the variables and to evaluate forecasts in terms of the variability of the cointegrating combinations of the errors, $\alpha' e_{t+h}$.

Accuracy measures based on cointegrating combinations of the forecast errors require that the cointegrating vector be known. Fortunately, such is often the case. The recent literature has emphasized repeatedly that economic and financial models frequently imply cointegration with simple and known cointegrating vectors (e.g., Watson 1994; Horvath and Watson 1995; Zivot 1996). Thus, although the assumption of a known cointegrating vector certainly involves a loss of generality, it is nevertheless legitimate in a variety of economically relevant cases. We will maintain the assumption of a known cointegration vector for the remainder of this article.

Interestingly, evaluation of accuracy in terms of the trace MSE of the cointegrating combinations of forecast errors is a special case of the general MSE measure. To see this, consider the general $N$-variate case with $r$ cointegrating relationships and consider again the MSE,

$$E(e_{t+h}' K e_{t+h}) = E\ \mathrm{trace}(e_{t+h}' K e_{t+h}) = E\ \mathrm{trace}(K e_{t+h} e_{t+h}') = \mathrm{trace}(K\Sigma_h),$$

where $\Sigma_h$ is the variance of $e_{t+h}$. Evaluating accuracy in terms of trace MSE of the cointegrating combinations of the forecast errors amounts to evaluating

$$E((\alpha' e_{t+h})'(\alpha' e_{t+h})) = \mathrm{trace}\ E((\alpha' e_{t+h})'(\alpha' e_{t+h})) = \mathrm{trace}(K\Sigma_h),$$

where $K = \alpha\alpha'$. Thus the trace MSE of the cointegrating combinations of the forecast errors is in fact a particular variant of MSE formulated on the raw forecast errors, $E(e'Ke) = \mathrm{trace}(K\Sigma_h)$, where the weighting matrix $K = \alpha\alpha'$ is of (deficient) rank $r(< N)$, the cointegrating rank of the system.

## 2.3 Accuracy Measures III: Trace MSE From the Triangular Representation

The problem with the traditional $E(e'Ke)$ approach with $K = I$ is that, although it values small MSE, it fails to value the long-run forecasts' hanging together correctly. Conversely, a problem with the $E(e'Ke)$ approach with $K = \alpha\alpha'$ is that it values *only* the long-run forecasts' hanging together correctly, whereas both pieces seem clearly relevant. The challenge is to incorporate both pieces into an overall accuracy measure in a natural way, and an attractive approach for doing so follows from the triangular representation of cointegrated systems exploited by Campbell and Shiller (1987) and Phillips (1991). Clements and Hendry (1995) provided a numerical example that illustrates the appeal of the triangular representation for forecasting. In the following, we provide a theoretical result that establishes the general validity of the triangular approach for distinguishing between naive univariate and fully specified system forecasts.

From the fact that $\alpha'$ has rank $r$, it is possible to rewrite the system so that the $N$ left-side variables are the $r$ error-correction terms followed by the differences of $N - r$ integrated but not cointegrated variables. That is, we rewrite the system in terms of

$$\begin{pmatrix} x_{1t} - \Gamma' x_{2t} \\ (1-L)x_{2t} \end{pmatrix},$$

where the variables have been rearranged and partitioned into $x_t = (x_{1t}', x_{2t}')'$, where $\Gamma = \Gamma(\alpha)$ and the variables in $x_{2t}$ are integrated but not cointegrated. We then evaluate accuracy in terms of the trace MSE of forecasts *from the triangular system,*

$$E\left\{\begin{pmatrix} e_{1,t+h} - \Gamma' e_{2,t+h} \\ (1-L)e_{2,t+h} \end{pmatrix}' \begin{pmatrix} e_{1,t+h} - \Gamma' e_{2,t+h} \\ (1-L)e_{2,t+h} \end{pmatrix}\right\}$$

$$= E\left\{\left(\begin{bmatrix} I_r & -\Gamma' \\ 0 & (1-L) \end{bmatrix} e_{t+h}\right)' \left(\begin{bmatrix} I_r & -\Gamma' \\ 0 & (1-L) \end{bmatrix} e_{t+h}\right)\right\},$$

which we denote trace $\text{MSE}_{\text{tri}}$. Notice that the trace $\text{MSE}_{\text{tri}}$ accuracy measure is also of $E(e'Ke)$ form, with

$$K = K(L) = \begin{bmatrix} I_r & -\Gamma' \\ 0 & (1-L) \end{bmatrix}' \begin{bmatrix} I_r & -\Gamma' \\ 0 & (1-L) \end{bmatrix}.$$

Recall Proposition 1, which says that under trace MSE, long-horizon forecast accuracy from the cointegrated system is no better than that from univariate models. We now show that, under trace $\text{MSE}_{\text{tri}}$, long-horizon forecast accuracy from the cointegrated system is *always better* than that from univariate models.

*Proposition 3.*

$$\lim_{h\to\infty} \frac{\text{trace}\widetilde{\text{MSE}}_{\text{tri}}}{\text{trace}\widehat{\text{MSE}}_{\text{tri}}} > 1.$$

*Proof.* Consider a cointegrated system in triangular form—that is, a system such that $\alpha' = [I - \Gamma']$. We need to show that, for large $h$,

$$\sum_{i=1}^{r} \text{var}[\alpha_i'\hat{e}_{t+h}] + \sum_{j=r+1}^{N} \text{var}[(1-L)\hat{e}_{j,t+h}] < \sum_{i=1}^{r} \text{var}[\alpha_i'\tilde{e}_{t+h}] + \sum_{j=r+1}^{N} \text{var}[(1-L)\tilde{e}_{j,t+h}]$$

and

$$\sum_{i=1}^{r} \text{var}[\alpha_i'\hat{e}_{t+h}] + \sum_{j=r+1}^{N} \text{var}[(1-L)\hat{e}_{j,t+h}] < \infty.$$

To establish the first inequality, it is sufficient to show that

$$\sum_{i=1}^{r} \text{var}[\alpha_i'\hat{e}_{t+h}] < \sum_{i=1}^{r} \text{var}[\alpha_i'\tilde{e}_{t+h}].$$

We showed earlier that, for large $h$,

$$\begin{aligned} \text{var}(\alpha'\tilde{e}_{t+h}) &= \alpha'Q\alpha + \alpha'\text{var}(C^*(L)\varepsilon_t + \Theta(L)u_t)\alpha \\ &\equiv \alpha'(Q+S)\alpha, \end{aligned}$$

where $Q \equiv \text{var}(\hat{e}_{t+h}), S \equiv \text{var}(C^*(L)\varepsilon_t + \Theta(L)u_t)$, and from which it follows that

$$\sum_{i=1}^{r} \text{var}[\alpha_i'\tilde{e}_{t+h}] - \sum_{i=1}^{r} \text{var}[\alpha_i'\hat{e}_{t+h}] = \text{trace}(\alpha'S\alpha) > 0$$

because $S$ is positive definite. To establish the second inequality, recall that

$$\hat{e}_{t+h} - \hat{e}_{t+h-1} = \sum_{i=1}^{h} C_{h-i}\varepsilon_{t+i} = C(L)\varepsilon_{t+h}$$

so that

$$\begin{aligned} &\text{var}[(1-L)\hat{e}_{t+h}] \\ &= \left(\sum_{j=0}^{h-1} C_j\right) \Omega \left(\sum_{j=0}^{h-1} C_j'\right) \to C(1)\Omega C(1)' \quad \text{as } h \to \infty. \end{aligned}$$

Let $C_{N-r}(1)$ be the last $N-r$ rows of $C(1)$; then altogether we have

$$\begin{aligned} &\sum_{i=1}^{r} \text{var}[\alpha_i'\hat{e}_{t+h}] + \sum_{j=r+1}^{N} \text{var}[(1-L)\hat{e}_{j,t+h}] \\ &\quad = \text{trace}(\alpha'Q\alpha) + \text{trace}(C_{N-r}(1)\Omega C_{N-r}(1)') < \infty, \end{aligned}$$

and the proof is complete.

In summary, although the long-horizon performances of the system and univariate forecasts are identical under the conventional trace MSE ratio, they differ under the trace $\text{MSE}_{\text{tri}}$ ratio. The system forecast is superior to the univariate forecast under trace $\text{MSE}_{\text{tri}}$ because the system forecast is accurate in the conventional "small MSE" sense *and* it makes full use of the information in the cointegrating relationship(s). We stress, however, that abandoning MSE and adopting $\text{MSE}_{\text{tri}}$ marks a change of loss function, and thus preferences. If the forecaster's loss function truly is trace MSE, then using trace $\text{MSE}_{\text{tri}}$ might not make sense. On the other hand trace MSE is often adopted without much thought, and an underlying theme of our analysis is precisely that thought *should* be given to the choice of loss function.

## 3. UNDERSTANDING EARLIER MONTE CARLO STUDIES

Here we clarify the interpretation of earlier influential Monte Carlo work, in particular Engle and Yoo (1987), as well as Reinsel and Ahn (1992), Clements and Hendry (1993), and Lin and Tsay (1996), among others. We do so by performing a Monte Carlo analysis of our own, which reconciles our theoretical results and the apparently conflicting Monte Carlo results reported in the literature, and we show how the existing Monte Carlo analyses have been misinterpreted. Throughout, we use a simple bivariate cointegrated system:

$$\begin{aligned} x_t &= \mu + x_{t-1} + \varepsilon_t \\ y_t &= \lambda x_t + v_t, \end{aligned}$$

where the disturbances are orthogonal at all leads and lags. The moving average representation is

$$\begin{aligned} (1-L)\begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} \mu \\ \lambda\mu \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \lambda & 1-L \end{pmatrix}\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \\ &\equiv \begin{pmatrix} \mu \\ \lambda\mu \end{pmatrix} + C(L)\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix}, \end{aligned}$$

and the error-correction representation is

$$\begin{aligned} (1-L)\begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} \mu \\ \lambda\mu \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 \\ 1 \end{pmatrix}(\lambda - 1)\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \lambda\varepsilon_t + v_t \end{pmatrix}. \end{aligned}$$

We set to $\lambda = 1, \mu = 0$, and $\sigma_\varepsilon^2 = \sigma_v^2 = 1$; we use a sample size of 100; and we perform 4,000 Monte Carlo replications.

This simple design allows us to make our point forcefully and with a minimum of clutter.

First consider Wiener–Kolmogorov forecasting from the multivariate cointegrated system. Write the time $t+h$ values in terms of time $t$ values and future innovations as

$$\begin{aligned} x_{t+h} &= \mu h + x_t + \sum_{i=1}^{h} \varepsilon_{t+i} \\ y_{t+h} &= \lambda(\mu h + x_t) + \lambda \sum_{i=1}^{h} \varepsilon_{t+i} + v_{t+h}, \end{aligned}$$

which makes clear that the $h$-step-ahead forecasts are

$$\begin{aligned} \hat{x}_{t+h} &= \mu h + x_t \\ \hat{y}_{t+h} &= \lambda \mu h + \lambda x_t. \end{aligned}$$

Now consider forecasting from the implied univariate representations; this, of course, first requires *finding* the univariate representations. The univariate representation for $x$ is of course a random walk with drift, exactly as given in the first equation of the system, $x_t = \mu + x_{t-1} + \varepsilon_t$. Hence the univariate forecast for $x$ is the same as the system forecast, $\tilde{x}_{t+h} = \mu h + x_t$. Derivation of the univariate representation for $y$ is a bit more involved. From the moving average representation of the system, rewrite the process for $y_t$ as a univariate two-shock process, $y_t = \lambda\mu + y_{t-1} + z_t$, where $z_t \equiv (1-L)v_t + \lambda\varepsilon_t$. Simple but tedious algebra reveals that $z_t$ is an MA(1) process, $z_t = u_t + \theta u_{t-1}$, where $\theta$ depends on the underlying parameters $\lambda, \sigma_\varepsilon^2$, and $\sigma_v^2$. [See Christoffersen and Diebold (1997a) for details of this and related calculations that appear in this section.] Thus $y$ is an integrated moving average of order 1. To form the univariate forecast for $y$, write

$$\begin{aligned} y_{t+h} &= \lambda\mu h + y_t + \sum_{i=1}^{h} z_{t+i} \\ &= \lambda\mu h + y_t + \theta u_t + u_{t+1} + \sum_{i=2}^{h} z_{t+i}, \end{aligned}$$

which makes clear that the time $t$ forecast for period $t+h$ is $\tilde{y}_{t+h} = \lambda\mu h + y_t + \theta u_t$.

Now let us proceed to the Monte Carlo analysis. The Wiener–Kolmogorov forecasts derived previously assume known parameters, but in practice the parameters must be estimated. Hence we replace population parameters with estimates to construct operational forecasts at each Monte Carlo replication, with the exception of the cointegrating vector, which, following the discussion of Section 2, is assumed known. In particular, for the univariate forecast we estimate $\theta$, but for the multivariate forecast we assume $\lambda$ is known. In Figure 1 we plot the trace MSE ratio and the trace $\text{MSE}_{\text{tri}}$ ratio against the forecast horizon, $h$. Using estimated parameters changes none of our theoretical results reached earlier under the assumption of known parameters. In particular, use of the trace MSE ratio obscures the long-horizon benefits of imposing cointegration, whereas use of trace $\text{MSE}_{\text{tri}}$ reveals those benefits clearly.

How then can we reconcile our results with those of Engle and Yoo (1987) and the many subsequent authors who

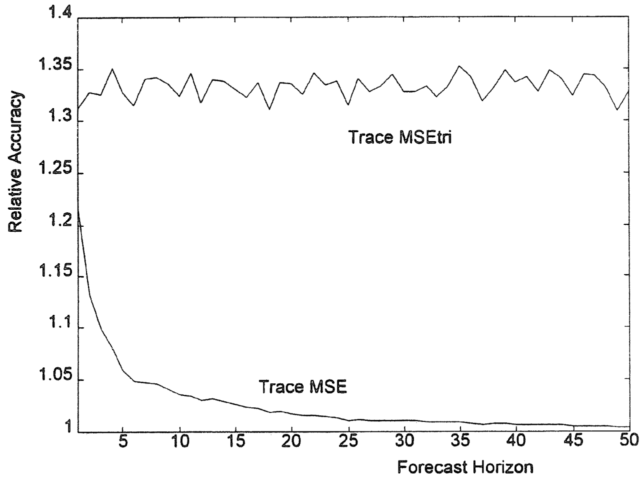


*Figure 1. Trace MSE Ratio and Trace $MSE_{tri}$ Ratio of Univariate Versus System Forecasts Plotted Against Forecast Horizon, Bivariate System With Estimated Parameters.*

concluded that imposing cointegration produces superior long-horizon forecasts? The answer is two-part—that literature makes a different and harder-to-interpret comparison than we do and it misinterprets the outcomes of the Monte Carlo experiments.

First consider the forecast comparison. We have thus far compared forecasts from univariate models (which impose integration) to forecasts from the cointegrated system (which impose both integration and cointegration). Thus, a comparison of the forecasting results isolates the effects of imposing cointegration. Engle and Yoo et al., in contrast, compare forecasts from a VAR in levels (which impose *neither* integration nor cointegration) to forecasts from the cointegrated system (which impose *both* integration and cointegration). Thus, differences in forecasting performance in the Engle–Yoo et al. setup cannot necessarily be attributed to the imposition of cointegration—instead, they may simply be due to imposition of integration, irrespective of whether cointegration is imposed.

Now consider the interpretation of the results. The VAR in levels is, of course, integrated, but estimating the system in levels entails estimating the unit root. Although many estimators are consistent, an exact finite-sample unit root is a zero-probability event. Unfortunately, even a slight and inevitable deviation of the estimated root from unity pollutes forecasts from the estimated model, and the pollution increases with $h$. This in turn causes the MSE *ratio* to increase in $h$ when comparing a levels VAR forecast to a system forecast or any other forecast that explicitly imposes unit roots. The problem is exacerbated by bias of the Dickey–Fuller–Hurwicz type; see Stine and Shaman (1989), Pope (1990), Abadir (1993), and Abadir, Hadri and Tzavalis (1996) for detailed treatments.

It is no surprise that forecasts from the VAR estimated in levels perform poorly, with performance worsening with horizon, as shown in Figure 2. It is tempting to attribute the poor performance of the VAR in levels to its failure to impose cointegration, as do Engle and Yoo et al. The fact

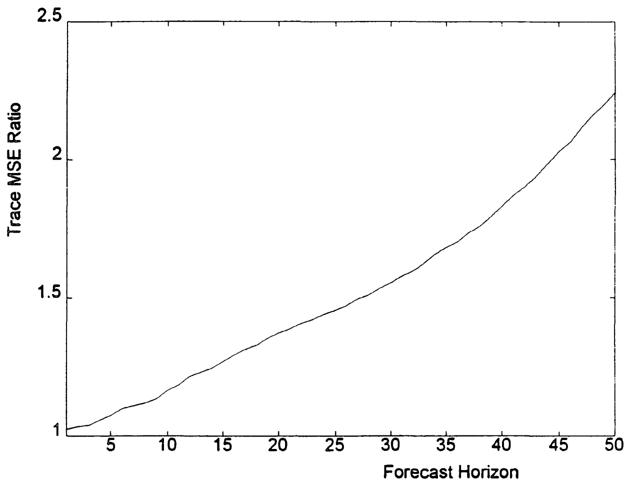


*Figure 2. Trace MSE Ratio of Levels VAR Versus Cointegrated System Forecasts Plotted Against the Forecast Horizon, Bivariate System With Estimated Parameters.*

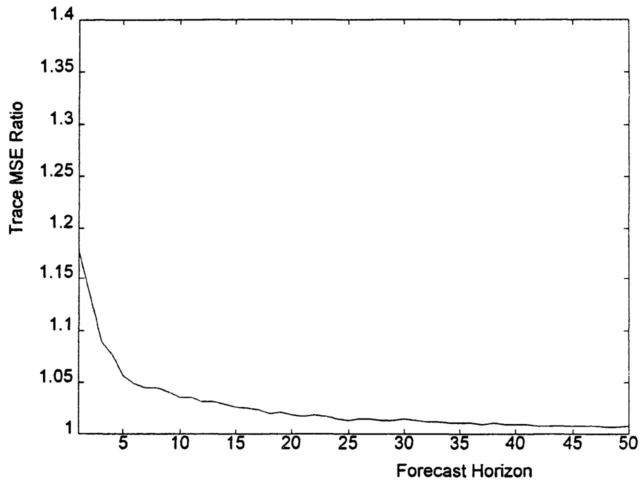


*Figure 3. Trace MSE Ratio of Differenced VAR Versus Cointegrated System Forecasts Plotted Against Forecast Horizon, Bivariate System With Estimated Parameters.*

is, however, that the VAR in levels performs poorly because it fails to impose *integration,* not because it fails to impose *cointegration*—estimation of the cointegrated system simply imposes the correct level of integration a priori. To see this, consider Figure 3, in which we compare the forecasts from an estimated VAR in *differences* to the forecasts from the estimated cointegrated system. At long horizons, the forecasts from the VAR in differences, which impose integration but completely ignore cointegration, perform just as well. In contrast, if we instead evaluate forecast accuracy with the trace $\mathrm{MSE}_{\mathrm{tri}}$ ratio that we have advocated, the forecasts from the VAR in differences compare poorly at all horizons to those from the cointegrated system, as shown in Figure 4.

In the simple bivariate system, we are restricted to studying models with exactly one unit root and one cointegration relationship. It is also of interest to examine richer systems; conveniently, the literature already contains relevant (but unnoticed) evidence, which is entirely consistent with our theoretical results. Reinsel and Ahn (1992) and Lin and Tsay (1996), in particular, provided Monte Carlo evidence on the comparative forecasting performance of competing estimated models. Both studied a four-variable VAR(2), with two unit roots and two cointegrating relationships. Their results clearly suggest that, under the trace MSE accuracy measure, one need only worry about imposing enough unit roots on the system. Imposing three (one too many) unit roots is harmless at *any* horizon, and imposing four unit roots (two too many so that the VAR is in differences) is harmless at long horizons. As long as one imposes enough unit roots, at least two in this case, the trace MSE ratio will invariably go to 1 as the horizon increases.

## 4. SUMMARY AND CONCLUDING REMARKS

First, we have shown that imposing cointegration does not improve long-horizon forecast accuracy when forecasts of cointegrated variables are evaluated using the standard trace MSE ratio. Ironically enough, although cointegration implies restrictions on low-frequency dynamics, imposing cointegration is helpful for short- but not long-horizon forecasting, in contrast to the impression created in the literature. Imposition of cointegration on an estimated system, when the system is in fact cointegrated, helps the accuracy of long-horizon forecasts relative to those from systems estimated in levels with no restrictions, but that is because of the imposition of integration, not cointegration. Univariate forecasts in differences do just as well! We hasten to add, of course, that the result is conditional on the assumption that the univariate representations of all variables do in fact contain unit roots. Differencing a stationary variable with roots close to unity has potentially dire consequences for long-horizon forecasting, as argued forcefully by Lin and Tsay (1996).

Second, we have shown that the variance of the cointegrating combination of the long-horizon forecast errors is

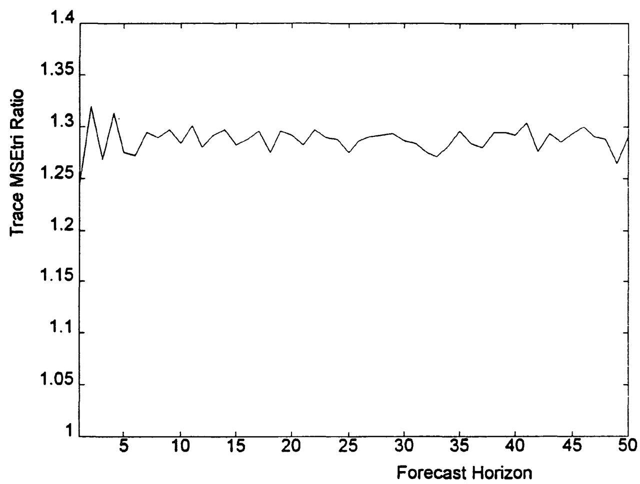


*Figure 4. Trace $MSE_{tri}$ Ratio of Differenced VAR Versus Cointegrated System Forecasts Plotted Against Forecast Horizon, Bivariate System With Estimated Parameters.*