Long-run forecasting in multi- and polynomially cointegrating systems.

Boriss Siliverstovs[§]
Centre for Dynamic Modelling in Economics
Department of Economics
University of Aarhus

Tom Engsted[¶]
Department of Finance
The Aarhus School of Business

Niels Haldrup[‖]
Centre for Dynamic Modelling in Economics
Department of Economics
University of Aarhus

December 22, 2001

Abstract

In this paper long-run forecasting in multi- and polynomially cointegrated models is investigated. It is shown that the result of Christoffersen and Diebold (1998) derived for I(1) cointegrating models generalizes to multi- and polynomially cointegrating systems. That is, in terms of the trace mean squared forecast error criterion, imposing the multi- and polynomially cointegrating restrictions does not lead to improved long-run horizon forecast accuracy when compared to forecasts generated from the univariate representations of the system variables. However, when one employs a loss function derived from the triangular representations of the (polynomially-) cointegrating systems, gains in forecastability are achieved for system forecasts as opposed to the univariate forecasts. The paper highlights the importance of carefully selecting loss functions in forecast evaluation of cointegrating systems.

Keywords: Forecasting, VAR models, Multicointegration, Polynomial Cointegration.

JEL Classification Codes: C32, C53.

[§]University of Aarhus, Department of Economics, Building 326, office 125, DK-8000 Aarhus C, Denmark, email: bsiliverstovs@econ.au.dk, (corresponding author)

[¶]The Aarhus School of Business, Department of Finance, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark, email: tom@asb.dk

[‖]University of Aarhus, Department of Economics, Building 326, office 217, DK-8000 Aarhus C, Denmark, email: nhaldrup@econ.au.dk

1 Introduction

Assessing the forecasting performance of econometric models is an important ingredient in evaluating such models. In multivariate models containing non-stationary variables, cointegration may be thought to play a key role in assessing forecasting ability, especially over long horizons, because cointegration captures the long-run comovement of the variables. Several studies have investigated the forecasting properties of cointegrated models. Engle and Yoo (1987) make a small Monte Carlo study where they compare mean-squared forecast errors from a VAR in levels, which imposes no cointegration, to forecasts from a correctly specified error-correction model (ECM), which imposes cointegration, and they find that longer-run forecasts from the ECM are more accurate. This result supports the above intuition that imposing cointegration gives better long-horizon forecasts for variables that are tied together in the long run. However, subsequent research has somewhat questioned and subsequently modified this – at the first glance appealing – conclusion.

According to Christoffersen and Diebold (1998), the doubts on the usefulness of cointegrating restrictions on the long-run forecasts are related to the following conjecture. The improved predictive power of cointegrating systems comes from the fact that deviations from the cointegrating relations tend to be eliminated. Thus, these deviations contain useful information on the likely future evolution of the cointegrating system which can be exploited to produce superior forecasts when compared to those made from models that omit cointegrating restrictions. However, since the long-run forecast of the cointegrating term is always zero, this useful information is only likely to be effective when the focus lies on producing the short-run forecasts. Hence, at least from this point of view, usefulness of imposing cointegrating relations for producing long-run superior forecasts can be questioned.

Clements and Hendry (1995) compare mean-squared error forecasts from a correctly specified ECM to forecasts from both an unrestricted VAR in levels and a misspecified VAR in first-differences (DVAR) which omits cointegrating restrictions present among the variables. They find that the forecasting superiority of the model that correctly imposes these cointegrating restrictions hinges crucially on whether the forecasts are for the levels of the variables, their first-differences, or the cointegrating relationship between the variables. They show that this difference in rankings for alternative yet isomorphic representations of the variables is due to the mean-squared forecast error (MSFE) criterion not being invariant to nonsingular, scale-preserving linear transformations of the model.[1] In particular, they show that the forecasts from the ECM model are not superior to those made from the DVAR model at all but the shortest forecast horizons when the first-differences of I(1) variables were forecasted.

Christoffersen and Diebold (1998) compare mean-squared error forecasts of the levels of I(1) variables

[1] Clements and Hendry (1993) suggest an alternative measure that is invariant to scale-preserving linear transformations of the data: the *generalized forecast error second moment* (GFESM) measure.

from a true cointegrated VAR to forecasts from correctly specified univariate representations, and they similarly find that imposing cointegration does not improve long-horizon forecast accuracy. Thus, it appears that the simple univariate forecasts are just as accurate as the multivariate forecasts when judged using the loss function based on the MSFE. They argue that this apparent paradox is due to the fact that the standard MSFE criterion fails to value the long-run forecasts' hanging together correctly. Long-horizon forecasts from the cointegrated VAR always satisfy the cointegrating restrictions exactly, whereas the long-horizon forecasts from the univariate models do so only on average, but this distinction is ignored in the MSFE criterion. Christoffersen and Diebold suggest an alternative criterion that explicitly accounts for this feature. The criterion is based on the *triangular representation* of cointegrated systems (see Campbell and Shiller, 1987, and Phillips, 1991). The virtue of this criterion is that it assesses forecast accuracy in the conventional "small MSFE" sense, but at the same time it makes full use of the information in the cointegrating relationships amongst the variables. Using this new forecast criterion, they indeed find that at long horizons the forecasts from the cointegrated VAR are superior to the univariate forecasts.

The purpose of the present paper is twofold. First of all, we extend the analysis of Christoffersen and Diebold to the case where the variables under study not only obey cointegrating relationships, but also obey certain *multi*cointegrating restrictions. Multicointegration was originally defined by Granger and Lee (1989, 1991) and refers to the case where the underlying I(1) variables are cointegrated in the usual sense *and* where, in addition, the cumulated cointegration errors cointegrate with the original I(1) variables. Thus, essentially there are two levels of cointegration amongst the variables. Multicointegration is a very convenient way of modeling the interactions between stock and flow variables. Granger and Lee considers the case where the two I(1) variables: production, y_t, and sales, x_t, cointegrate, such that inventory investments, z_t, are stationary, $z_t \equiv y_t - \beta x_t \sim \mathrm{I}(0)$, but where the cumulation of inventory investment, $I_t \equiv \Sigma_{j=1}^{t} z_j$, i.e. the level of inventories (which is then an I(1) stock variable), in turn cointegrates with either y_t or x_t, or both of them. Another example, analyzed by Lee (1992) and Engsted and Haldrup (1999), is where y_t is new housing units started, x_t is new housing units completed, z_t is uncompleted starts, and hence I_t is housing units under construction. Leachman (1996), and Leachman and Francis (2000) provide examples of multicointegrating systems with government revenues and expenditures, and a country's export and import, respectively. Here the stock variable is defined as the government debt (surplus) and the country's external debt (surplus), such that each variable is the cumulated series of past government and trade deficits, respectively. Yet another example is provided by Siliverstovs (2001) who analyze consumption and income, and where cumulated savings (i.e. the cumulation of the cointegrating relationship between income and consumption) constitutes wealth, which cointegrates with consumption and income. In general, multicointegration captures the notion of *integral control* in dynamic systems, see, for example, Hendry and von Ungern Sternberg (1981) *inter alia*.

Testing for multicointegration, and estimation of models with multicointegrating restrictions, are most naturally conducted within an I(2) cointegration framework, see Engsted, Gonzalo and Haldrup (1997), Haldrup (1999), Engsted and Johansen (1999), and Engsted and Haldrup (1999). In the present paper we investigate how the presence of multicointegration affects forecasting comparisons. In particular, we set up a model that contains both cointegrating and multicointegrating restrictions, and then we examine how forecasts from this multicointegrated system compares to univariate forecasts. The comparison is done in terms of (trace) mean-squared forecast errors, but we follow Christoffersen and Diebold (1998) in using both a standard loss function and a loss function based on the triangular representation of the cointegrating system.

Secondly, we extend the analysis of Christoffersen and Diebold (1998) to the case when forecasting is undertaken in so called polynomially cointegrating systems, where the original I(2) variables cointegrate with their first differences[2]. Hence, the forecasting of I(2) variables constitutes the primary interest in the polynomially cointegrating systems. Examples of polynomially contegrating systems have been given in Rahbek, Kongsted, and Jørgensen (1999), and Banerjee, Cockerell, and Russell (2001), *inter alia*. Rahbek et. al. (1999) find a polynomially cointegrating relation in a UK money demand data set, which involves both levels and first differences of the original I(2) variables represented by the logarithmic transformation of nominal money and nominal prices levels. Banerjee et. al. (2001) analyse the system of I(2) variables which consists of the nominal price level and unit labour and import costs, expressed in logarithms. They find the polynomially cointegrating relation between the markup – defined as a particular linear combination between the price level and costs – on the one hand, and the inflation rate, on the other hand. As in the section on forecasting in multicointegrating systems, we compare the forecasting performance of the model that imposes the polynomially cointegrating relations with the forecasts made from the correctly specified univariate models for the I(2) variables. We use the loss function based on the MSFE as well as the loss function based on the triangular representation of the polynomially cointegrating variables. The latter loss function takes into account the fact that the system long-run forecasts maintain the polynomially cointegrating relations exactly as opposed to the their univariate competitors, which satisfy the polynomially cointegrating relations only on average. The former loss function fails to acknowledge such a distinction between the system and univariate forecasts.

Our most important results can be summarized as follows. First, we find that the general result of Christoffersen and Diebold derived for the cointegrating models carries over to multicointegrated as well as polynomially cointegrated models. Based on the traditional MSFE criterion, long-horizon forecasts

[2] Observe that despite the fact that the inference and analysis of multicointegrated as well as polynomially cointegrated systems are technically the same, we choose to keep this distinction in terminology. We refer to the multicointegrating system when dealing with original I(1) variables and its respective cumulants. On the opposite, the polynomially cointegrated system is referred to when dealing with original I(2) variables and its first difference transformation.

from the multi- and polynomially cointegrated systems are found not to be superior to simple univariate forecasts. However, based on the triangular MSFE criterion the system forecasts are clearly superior to the univariate forecasts. Second, we find that if focus is on forecasting the original I(1) variables of the system, nothing is lost by ignoring the multicointegrating property of the system when evaluating its forecasting performance: although the forecasts are constructed from the correctly specified multicointegrated model, when *evaluating* these forecasts, one can just use Christoffersen and Diebold's triangular MSFE measure that includes the first layer of cointegration but excludes the second layer of cointegration represented by the multicointegrating relation. The explanation is that the long-run forecasts of the multicointegrating I(1) variables maintain the *cointegrating* but not the multicointegrating relations. Hence, it seems appropriate to evaluate the long-run forecasting performance of the multicointegrating system using the loss function of Christoffersen and Diebold, which in particular values maintainance of the cointegrating relations in the long-run. Third, if focus is on forecasting the I(2) variables of the system then forecasts should be based on the polynomially cointegrating system, *and* in evaluating the forecasts one should use an extended triangular MSFE criterion, which explicitly acknowledges the maintainance of the polynomially cointegrating restrictions amongst the long-run system forecasts.

Observe that due to the fact that our primary interest is on the particular dynamic characteristics of multi- and polynomial cointegration with respect to forecasting, we abstract from estimation issues and hence assume known parameters.

The rest of the paper is organized as follows. In Section 2 we set up the multi– and polynomially cointegrating systems used in the subsequent analysis. Also, we derive the corresponding univariate representations of the system variables. Sections 3 and 4 discuss the long-run forecasting in the multi– and polynomially cointegrating systems, respectively. Section 5 illustrates our findings using a numerical example and the final section concludes.

2 Multi- and polynomially cointegrating sytems.

In this section we define multicointegrating and polynomially cointegrating models and derive the corresponding univariate representations of the system variables. To ease the exhibition we employ the simplest models with relevant multi– and polynomially cointegrating restrictions.

2.1 Multicointegrating system.

First, we address a model incorporating multicointegrating restrictions. Consider the two I(1) variables, x_t and y_t, that obey a cointegrating relation

$$y_t - \lambda x_t \sim I(0), \tag{1}$$

such that the cumulated cointegration error

$$\sum_{j=1}^{t} (y_j - \lambda x_j) \sim I(1)$$

is an I(1) variable by construction[3]. We refer to the system as multicointegrating when there exists a stationary linear combination of the cumulated cointegrating error and the original variables, e.g.

$$\sum_{j=1}^{t} (y_j - \lambda x_j) - \alpha x_t \sim I(0). \tag{2}$$

As discussed in Granger and Lee (1989, 1991), the multicointegrating restrictions are likely to occur in stock-flow models, where both cointegrating relations have an appealing interpretation. The first cointegrating relation (1) is formed between the original flow variables, for example: production and sales, income and expenditures, export and import, etc. The second cointegrating relation (2) represents the relation between the cumulated past descrepancies between the flow variables, for instance: the stock of inventories, the stock of wealth, the stock of external debt (surplus), and all or some flow variables present in the system. It implies that the equilibrium path of the system is maintained not only through the flow variables alone, but there are additional forces tying together the stock and flow series and in so doing providing a second layer of equilibrium.

It is convenient to represent the system of the multicointegrating variables in the triangular form

$$\begin{aligned} \Delta x_t &= e_{1t} \\ \sum_{j=1}^{t} y_j &= \lambda \sum_{j=1}^{t} x_j + \alpha x_t + e_{2t}, \end{aligned} \tag{3}$$

where the cumulated I(1) series are now I(2), by construction. The disturbances are uncorrelated at all leads and lags, i.e. $E(e_{1t-j}e_{2t-i}) = 0$, $\forall\, j \neq i$ for $j = 0, \pm1, \pm2, ...$ and $i = 0, \pm1, \pm2, ...$. We denote the I(2) variables by capital letters, i.e. $Y_t = \sum_{j=1}^{t} y_j$ and $X_t = \sum_{j=1}^{t} x_j$.[4] This allows us to write the system

[3]Note that no deterministic components are assumed in the series and hence, by construction, no trend, for instance, are generated in the cumulated series.

[4]However, it is worthwhile keeping in mind the distinction bewteen I(2) variables in the multi- and polynomially cointegrating systems. In the former case, they are generated as such, whereas in the latter case the original series are I(2).

above as

$$\begin{aligned} \Delta x_t &= e_{1t} \\ Y_t &= \lambda X_t + \alpha x_t + e_{2t}. \end{aligned} \tag{4}$$

Below we provide two equivalent representations of the system in (4). The Vector Error-Correction model (**VECM**) can be represented as follows:

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} [y_{t-1} - \lambda x_{t-1}] + \begin{bmatrix} 0 \\ -1 \end{bmatrix} [Y_{t-1} - \lambda X_{t-1} - \alpha x_{t-1}] + \begin{bmatrix} e_{1t} \\ (\lambda + \alpha) e_{1t} + e_{2t} \end{bmatrix}.$$

As seen, the **VECM** explicitly incorporates both cointegration levels, see equations (1) and (2), that are present in the multicointegrating system (4). Alternatively, the multicointegrated system (4) can be given the **moving-average** representation:

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = C(L) e_t = \begin{bmatrix} 1 & 0 \\ [\lambda + (1 - L)\alpha] & (1 - L)^2 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}. \tag{5}$$

Granger and Lee (1991) argue that the necessary and sufficient condition for the time series x_t and y_t to be multicointegrated is that the determinant of $C(L)$ should have a root $(1 - L)^2$. This condition is clearly satisfied for our simple system.

2.2 Polynomially cointegrating system.

In contrast to the multicointegrating model where two equilibrium layers are of equal interest, there is one cointegrating relation that is of primary interest in the polynomially cointegrating systems. Namely, the one that is formed by the levels of the original I(2) variables and its first differences

$$Y_t = \lambda X_t + \alpha \Delta X_t + e_{2t}.$$

The most popular example of such a polynomially cointegrating relation involves levels of nominal monetary and/or price variables (expressed in logarithms) and inflation (defined as the first difference of the logarithmic transformation of the nominal prices), see Rahbek et.al. (1999), for example.

Similarly to the multicointegrating model, we can write the polynomially cointegrating model in the triangular form as follows

$$\begin{aligned} \Delta^2 X_t &= e_{1t} \\ Y_t &= \lambda X_t + \alpha \Delta X_t + e_{2t}, \end{aligned} \tag{6}$$

where it is assumed that the disturbances are orthogonal at all lags and leads. Observe that in this simple system the common I(2) trend is represented by X_t and the respective polynomially cointegrating relation is given by $Y_t - \lambda X_t - \alpha \Delta X_t = e_{2t}$.

The polynomially cointegrating system (6) can be given the **moving-average** representation:

$$\begin{bmatrix} \Delta^2 X_t \\ \Delta^2 Y_t \end{bmatrix} = C(L)\, e_t = \begin{bmatrix} 1 & 0 \\ [\lambda + (1-L)\,\alpha] & (1-L)^2 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}. \tag{7}$$

Observe that the representations (5) and (7) are equivalent, however the focus on I(1) and I(2) variables differs.

2.3 Univariate representations.

In this section we derive the implied univariate representations for the x_t and y_t as well as for X_t and Y_t series. For the original variable x_t and its cumulative counterpart X_t the univariate representations remain the same as given in the system (4)

$$\begin{aligned} x_t &= x_{t-1} + e_{1t} \\ X_t &= X_{t-1} + \Delta X_{t-1} + e_{1t}. \end{aligned}$$

In deriving the implied univariate representation for y_t and Y_t we follow Christoffersen and Diebold (1998) by matching the autocovariances of the process z_t. From the MA-representation of $\Delta^2 Y_t = \Delta y_t$ we have

$$\begin{aligned} \Delta^2 Y_t &= \Delta y_t = [\lambda + (1-L)\,\alpha]\, e_{1t} + (1-L)^2\, e_{2t} \\ y_t &= y_{t-1} + z_t \\ Y_t &= Y_{t-1} + \Delta Y_{t-1} + z_t, \end{aligned}$$

where, as shown in the appendix, the process z_t corresponds to the MA(2) process

$$z_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}, \quad u_t \sim \mathsf{IID}\left(0, \sigma_u^2\right). \tag{8}$$

The coefficient θ_2 represents a root of the following fourth order polynomial

$$\begin{gathered} \theta_2^4 + (2-B)\,\theta_2^3 + \left(A^2 - 2B + 2\right)\theta_2^2 + (2-B)\,\theta_2 + 1 = 0, \\ \text{where} \quad A = [-\alpha\,(\lambda+\alpha)\,q - 4]\,, \; B = \left[(\lambda+\alpha)^2 + \alpha^2\right] q + 6, \text{ and } q = \frac{\sigma_1^2}{\sigma_2^2}. \end{gathered} \tag{9}$$

and the coefficient θ_1 and the variance term σ_u^2 can be found as follows:

$$\theta_1 = \frac{\theta_2}{(1+\theta_2)} A \quad \text{and} \quad \sigma_u^2 = \frac{\sigma_2^2}{\theta_2}. \tag{10}$$

Observe that the values of the MA coefficients θ_1 and θ_2 are chosen such that they satisfy the invertibility conditions for the MA(2) process z_t.

3 Long-run forecasting in multicointegrating systems.

In this section we closely follow the approach of Christoffersen and Diebold (1998) in comparing the long-run forecasting performance of the model that correctly imposes multicointegration − and at the same time cointegration − restrictions and the univariate model that omits those restrictions completely. Hence, our analysis extends the results of Christoffersen and Diebold (1998) derived for cointegrated I(1) systems to multicointegrating systems where focus is on forecasting I(1) variables.[5]

In order to motivate the subsequent analysis of long-run forecasting in the multicointegrating systems, it is worthwhile reviewing related results of Christoffersen and Diebold (1998) for the long-run forecasts in standard I(1) cointegrating systems. Christoffersen and Diebold (1998) show that when comparing the forecasting performance of the models that impose cointegration and correctly specified univariate models in terms of MSFE, there are no gains of imposing cointegration at all but the shortest forecast horizons. The problem is that the MSFE criterion fails to acknowledge the important distinction between the long-run system forecasts and the univariate forecasts. That is, the intrinsic feature of the long-run system forecasts is that they preserve the cointegrating relations exactly, whereas the long-run forecasts from the univariate models satisfy the cointegrating relations on average only. As a result, the variance of a cointegrating combination of the system forecast errors will always be smaller than that of the univariate forecast errors.

Therefore, if one can define a loss function which recognizes this distinction between the system- and univariate forecasts, then it becomes possible to discriminate between the forecasts made from these models. Christoffersen and Diebold (1998) show that such a loss function can be based on the triangular representation of the cointegrating variables, see Campbell and Shiller (1987), and Phillips (1991). In its simplest form a standard I(1) cointegrated system reads

$$\begin{bmatrix} 1 & -\lambda \\ 0 & 1-L \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} v_{2t} \\ v_{1t} \end{bmatrix},$$

where it is assumed that the disturbance terms are uncorrelated at all leads and lags. The corresponding loss function, introduced in Christoffersen and Diebold (1998), looks as follows:

$$\text{trace MSFE}_{tri} = \mathsf{E}\left[\left(\begin{bmatrix} 1 & -\lambda \\ 0 & 1-L \end{bmatrix} \begin{bmatrix} v_{2t+h} \\ v_{1t+h} \end{bmatrix}\right)' \left(\begin{bmatrix} 1 & -\lambda \\ 0 & 1-L \end{bmatrix} \begin{bmatrix} v_{2t+h} \\ v_{1t+h} \end{bmatrix}\right)\right], \tag{11}$$

such that the forecast accuracy of a given model is judged upon the corresponding forecast errors v_{1t+h}

[5] In the subsequent anlysis the issue of estimation is being abstracted from. Estimation of unknown parameters is naturally of essential importance in forecasting. However, in order to address the forecast structure of multi- as well as polynomially cointegrating systems in particular, we assume that the parameters are known in the considered models.

and v_{2t+h}. The trace MSFE_{tri} also reads

$$\text{trace MSFE}_{tri} = \mathsf{E}\left[\begin{bmatrix} v_{2t+h} \\ v_{1t+h} \end{bmatrix}' K \begin{bmatrix} v_{2t+h} \\ v_{1t+h} \end{bmatrix}\right], \text{ where } K = \begin{bmatrix} 1 & -\lambda \\ 0 & 1-L \end{bmatrix}' \begin{bmatrix} 1 & -\lambda \\ 0 & 1-L \end{bmatrix}$$

and it is instructive to compare this with the traditional MSFE used in other studies:

$$\text{trace MSFE} = \mathsf{E}\left[\begin{bmatrix} v_{2t+h} \\ v_{1t+h} \end{bmatrix}' \begin{bmatrix} v_{2t+h} \\ v_{1t+h} \end{bmatrix}\right].$$

As seen, the traditional MSFE can be regarded as the special case of the trace MSFE_{tri} with $K = I$. The trace MSFE_{tri} criterion values the small forecasts errors as does the traditional MSFE criterion, but at the same time it also values maintainance of the cointegrating restrictions amongst the generated forecasts. With other things being equal, the latter fact proved to be cruical in distinguishing between the system- and univariate forecasts.

In the present section we extend the results of Christoffersen and Diebold (1998) to the multicointegrating model forecasts. As shown below, the basic structure of their argument carries over to the case of our interest. Firstly, the usual MSFE criterion fails to distinguish between the forecasts generated from the multicointegrating models and the corresponding univariate models at the long forecast horizons. Secondly, the multicointegrating system forecasts obey cointegrating but not multicointegrating relations in the long run, whereas the corresponding univariate forecasts maintain the cointegrating relations on average only. Thirdly, although the variance of the forecast errors of levels of I(1) variables grows of order $O(h)$, the variance of the cointegrating combination of these forecast errors is finite. This is relevant both for system- and univariate forecasts, however, in the latter case the variance of the cointegrating combination is greater than in the former case.

The message is that, although forecasts are made from the multicointegrating model, evaluation of the forecasts can be carried out by means of the loss function (11), which is based on the triangular representation of the cointegrating system. The rest of this section illustrates this important conclusion.

3.1 Forecasting I(1) variables from the multicointegrating system.

The MA-representation of the multicointegrating variables (5) allows us to write the evolution of the multicointegrating system in terms of time t values x_t and future innovations e_{1t+h} and e_{2t+h}:

$$\begin{aligned} x_{t+h} &= x_t + \sum_{i=1}^{h} e_{1t+i} \\ y_{t+h} &= \lambda x_t + \lambda \sum_{i=1}^{h} e_{1t+i} + \alpha e_{1t+h} + \Delta e_{2t+h} \end{aligned}$$

Correspondingly, the h-steps ahead forecasts for I(1) variables are given by

$$\begin{aligned} \widehat{x}_{t+h} &= x_t \\ \widehat{y}_{t+h} &= \lambda x_t \end{aligned} \tag{12}$$

for **all** forecast horizons but $h = 1$. In the latter case we have

$$\begin{aligned} \widehat{x}_{t+1} &= x_t \\ \widehat{y}_{t+1} &= \lambda x_t - e_{2t} = \lambda x_t - [Y_t - \lambda X_t - \alpha x_t] . \end{aligned} \tag{13}$$

In particular, observe the long-run forecasts from our multicointegrating system maintain the cointegrating relation exactly:

$$\widehat{y}_{t+h} = \lambda \widehat{x}_{t+h}. \tag{14}$$

Continuing, the forecast errors read

$$\begin{aligned} \widehat{\varepsilon}_{x,t+h} &= \sum_{i=1}^{h} e_{1t+i} \quad \forall h > 0 \\ \widehat{\varepsilon}_{y,t+h} &= \begin{cases} \lambda e_{1t+1} + \alpha e_{1t+1} + e_{2t+1} = (\lambda + \alpha) e_{1t+1} + e_{2t+1} & \text{for } h = 1 \\ \lambda \sum_{i=1}^{h} e_{1t+i} + \alpha e_{1t+h} + \Delta e_{2t+h} & \text{for } h > 1 \end{cases} \end{aligned}$$

Furthermore, we can note that the forecast errors and the original system as in (5) follow the same stochastic process, i.e.

$$\begin{bmatrix} \Delta \widehat{\varepsilon}_{x,t+h} \\ \Delta \widehat{\varepsilon}_{y,t+h} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda + \alpha (1 - L) & (1 - L)^2 \end{bmatrix} \begin{bmatrix} e_{1t+h} \\ e_{2t+h} \end{bmatrix}. \tag{15}$$

The forecast error variances are thus given by

$$Var (\widehat{\varepsilon}_{x,t+h}) = h\sigma_1^2 \sim O (h) , \text{ for } h > 0 \tag{16}$$

$$Var (\widehat{\varepsilon}_{y,t+h}) = \begin{cases} (\lambda + \alpha)^2 \sigma_1^2 + \sigma_2^2 & \text{for } h = 1 \\ \lambda^2 \sigma_1^2 h + \left[(\lambda + \alpha)^2 - \lambda^2 \right] \sigma_1^2 + 2\sigma_2^2 \sim O (h) & \text{for } h > 1 \end{cases} \tag{17}$$

Notice that the variance of the system forecast error for y_{t+h} and x_{t+h} is growing of order $O (h)$. The variance of the cointegrating combination of the forecast errors is given by

$$Var (\widehat{\varepsilon}_{y,t+h} - \lambda \widehat{\varepsilon}_{x,t+h}) = \begin{cases} \alpha^2 \sigma_1^2 + \sigma_2^2 & \text{for } h = 1 \\ \alpha^2 \sigma_1^2 + 2\sigma_2^2 & \text{for } h > 1 \end{cases} \tag{18}$$

and hence is finite for all forecast horizons. Observe that in this simple model the variance of the cointegrating combination of the forecast errors is the same for all forecast horizons except for $h = 1$. The reason for the latter finding can be seen from equations (12) and (13), which reflects the fact that the multicointegrating term is in the information set for $h = 1$ and it has expectation zero for $h > 1$.

SB

3.2 Forecasts from implied univariate representations for I(1) variables.

Next, we turn to the forecasting of the I(1) variables based on the correctly specified implied univariate representations. Future values of x_{t+h} are given by

$$x_{t+h} = x_t + \sum_{i=1}^{h} e_{1t+i}$$

and for the variable y_{t+h}

$$y_{t+h} = \begin{cases} y_t + z_{t+1} = y_t + u_{t+1} + \theta_1 u_t + \theta_2 u_{t-1} & h = 1 \\ y_t + \sum_{i=1}^{h} z_{t+i} = y_t + u_{t+1} + \theta_1 u_t + \theta_2 u_{t-1} + u_{t+2} + \theta_1 u_{t+1} + \theta_2 u_t + \sum_{i=3}^{h} z_{t+i} & h > 1 \end{cases}$$

The corresponding $h-$steps ahead forecasts for I(1) variables can now be derived as follows. The forecast for the variable x_t is the same as the system forecast

$$\widetilde{x}_{t+h} = \widehat{x}_{t+h} = x_t$$

whereas the forecast $\widetilde{y}_{t+h}$ is given by

$$\widetilde{y}_{t+h} = \begin{cases} y_t + \theta_1 u_t + \theta_2 u_{t-1} & \text{for } h = 1 \\ y_t + \theta_1 u_t + \theta_2 u_{t-1} + \theta_2 u_t = y_t + (\theta_1 + \theta_2) u_t + \theta_2 u_{t-1}, & \text{for } h > 1 \end{cases}$$

The forecast error and the corresponding forecast error variance for x_{t+h} are given by

$$\begin{aligned} \widetilde{\varepsilon}_{x,t+h} &= \widehat{\varepsilon}_{x,t+h} = \sum_{i=1}^{h} e_{1t+i} \\ Var\left(\widetilde{\varepsilon}_{x,t+h}\right) &= Var\left(\widehat{\varepsilon}_{x,t+h}\right) = h\sigma_1^2 \sim O\left(h\right). \end{aligned} \tag{19}$$

The corresponding forecast error $\widetilde{\varepsilon}_{y,t+h} = y_{t+h} - \widetilde{y}_{t+h}$ for the variable y_t reads

$$\widetilde{\varepsilon}_{y,t+h} = \begin{cases} u_{t+1} & \text{for } h = 1 \\ u_{t+1} + u_{t+2} + \theta_1 u_{t+1} + \sum_{i=3}^{h} z_{t+i} = (1 + \theta_1 + \theta_2) \sum_{i=1}^{h-2} u_{t+i} + (1 + \theta_1) u_{t+h-1} + u_{t+h} & \text{for } h > 1 \end{cases}$$

with the forecast variance

$$Var(\widetilde{\varepsilon}_{y,t+h}) = \begin{cases} \sigma_u^2 & \text{for } h = 1 \\ \left[(1 + \theta_1 + \theta_2)^2 (h-2) + (1 + \theta_1)^2 + 1\right] \sigma_u^2 = & \\ = \lambda^2 \sigma_1^2 (h-2) + \left[(1 + \theta_1)^2 + 1\right] \sigma_u^2 \sim O\left(h\right) & \text{for } h > 1. \end{cases} \tag{20}$$

Next we derive the variance of the cointegrating combination of the forecast errors:

$$Var\left(\widetilde{\varepsilon}_{y,t+h} - \lambda \widetilde{\varepsilon}_{x,t+h}\right) = Var\left(\widetilde{\varepsilon}_{y,t+h}\right) + \lambda^2 Var\left(\widetilde{\varepsilon}_{x,t+h}\right) - 2\lambda cov(\widetilde{\varepsilon}_{y,t+h}, \widetilde{\varepsilon}_{x,t+h}),$$

using expressions (19) and (20) and the following expression for the covariance term

$$cov(\widetilde{\varepsilon}_{y,t+h}, \widetilde{\varepsilon}_{x,t+h}) = \lambda h \sigma_1^2 + \alpha \sigma_1^2.$$

The variance of the cointegrating combination of the forecast errors reads

$$Var\left(\widetilde{\varepsilon}_{y,t+h} - \lambda \widetilde{\varepsilon}_{x,t+h}\right) = \begin{cases} \sigma_u^2 - \lambda^2 \sigma_1^2 - 2\lambda\alpha\sigma_1^2 & \text{for } h = 1 \\ -2\lambda^2\sigma_1^2 - 2\lambda\alpha\sigma_1^2 + \left[(1+\theta_1)^2 + 1\right]\sigma_u^2 < \infty & \text{for } h > 1 \end{cases}. \quad (21)$$

This implies that the variance of the cointegrating combination of forecasts from the implied univariate representations is *finite*.

3.3 Comparison of the forecast accuracy for I(1) variables.

First we use the trace MSFE criterion to compare the forecast accuracy of the multivariate and univariate forecast representations. Using equations (16) and (17) , and (19) and (20) we can calculate the behaviour of the conventional measure of the forecast accuracy (trace MSFE) as the forecast horizon increases:

$$\frac{trace(var(\widetilde{\varepsilon}_{t+h}))}{trace(var(\widehat{\varepsilon}_{t+h}))} = \frac{h\sigma_1^2 + \lambda^2 (h-1)\sigma_1^2 - \lambda^2\sigma_1^2 + \left[(1+\theta_1)^2 + 1\right]\sigma_u^2}{h\sigma_1^2 + \lambda^2 (h-1)\sigma_1^2 + (\lambda+\alpha)^2\sigma_1^2 + 2\sigma_2^2} \to 1 \quad (22)$$

As seen, as $h \to \infty$ this trace ratio approaches 1 since the coefficients to the leading terms both in nominator and denominator are identical. That is, on the basis of the traditional forecast evaluation criterion (trace MSFE) it is impossible to distinguish between the model with imposed multicointegration restrictions and the model that ignores these restrictions completely. The well-recognized drawback of the trace MSFE criterion is that it fails to value the exact maintainance of cointegrating relations by the long-run forecasts. Hence, the solution is to employ the loss function that recognizes this fact. Recall that we have shown above that the long-run forecasts from the multicointegrating system obey the cointegrating relation exactly. Therefore it seems natural to adopt the loss function based on the triangular system for the cointegrating variables for our purposes. Using the loss function (11) as defined in Christoffersen and Diebold (1998), we have for the system forecasts

$$trace\ \widehat{\text{MSFE}}_{tri} = E\left\{ \begin{pmatrix} \widehat{\varepsilon}_{y,t+h} - \lambda\widehat{\varepsilon}_{x,t+h} \\ (1-L)\,\widehat{\varepsilon}_{x,t+h} \end{pmatrix}' \begin{pmatrix} \widehat{\varepsilon}_{y,t+h} - \lambda\widehat{\varepsilon}_{x,t+h} \\ (1-L)\,\widehat{\varepsilon}_{x,t+h} \end{pmatrix} \right\}.$$

That is

$$trace\ \widehat{\text{MSFE}}_{tri} = \begin{cases} \alpha^2\sigma_1^2 + \sigma_2^2 + \sigma_1^2 & \text{for } h = 1 \\ \alpha^2\sigma_1^2 + 2\sigma_2^2 + \sigma_1^2 & \text{for } h > 1 \end{cases}.$$

For the forecasts from the univariate models we have

$$trace\ \widetilde{\text{MSFE}}_{tri} = E\left\{ \begin{pmatrix} \widetilde{\varepsilon}_{y,t+h} - \lambda\widetilde{\varepsilon}_{x,t+h} \\ (1-L)\,\widetilde{\varepsilon}_{x,t+h} \end{pmatrix}' \begin{pmatrix} \widetilde{\varepsilon}_{y,t+h} - \lambda\widetilde{\varepsilon}_{x,t+h} \\ (1-L)\,\widetilde{\varepsilon}_{x,t+h} \end{pmatrix} \right\}$$

and thus

$$trace\ \widetilde{\text{MSFE}}_{tri} = \begin{cases} \sigma_u^2 - \lambda^2\sigma_1^2 - 2\lambda\alpha\sigma_1^2 + \sigma_1^2 & \text{for } h = 1 \\ \left[(1+\theta_1)^2 + 1\right]\sigma_u^2 - 2\lambda^2\sigma_1^2 - 2\lambda\alpha\sigma_1^2 + \sigma_1^2 & \text{for } h > 1 \end{cases}.$$

Comparing the ratios we have

$$\frac{trace\ \widetilde{\text{MSFE}}_{tri}^{h>1}}{trace\ \widehat{\text{MSFE}}_{tri}^{h>1}} = \frac{\left[(1+\theta_1)^2 + 1\right]\sigma_u^2 - 2\lambda^2\sigma_1^2 - 2\lambda\alpha\sigma_1^2 + \sigma_1^2}{\alpha^2\sigma_1^2 + 2\sigma_2^2 + \sigma_1^2} > 1 \tag{23}$$

$$\frac{trace\ \widetilde{\text{MSFE}}_{tri}^{h=1}}{trace\ \widehat{\text{MSFE}}_{tri}^{h=1}} = \frac{\sigma_u^2 - \lambda^2\sigma_1^2 - 2\lambda\alpha\sigma_1^2 + \sigma_1^2}{\alpha^2\sigma_1^2 + \sigma_2^2 + \sigma_1^2} > 1. \tag{24}$$

It is not straightforward to show analyticaly that the above inequalities apply. However, using numerical simulation it can be shown that the trace ratios (23) and (24) will always be greater than unity.[6]

In summary, a number of results of Christoffersen and Diebold (1998) derived for the cointegrating systems straightforwardly carries out to the model that obeys multicointegrating restrictions. First, long-run forecasts generated from the multicointegrating system preserve the cointegrating relations exactly, see (14). Second, the system forecast errors follow the same stochastic process as the original variables, as depicted in (15). Third, the variance of the cointegrating combination of the system forecast errors is finite (see (18)) even though the variance of the system forecast errors of a separate variable grows of order $O(h)$, as seen in expression (17). Fourth, the variance of the cointegrating combination of the univariate forecast errors is finite too even so the variance of the univariate forecast errors grows of order $O(h)$, see (21), (19), and (20). Fifth, imposing the multicointegrating restrictions does not lead to the improved forecast performance over the univariate models when compared in terms of the traditional mean squared forecast error criterion, as shown in (22). Finally, adoption of the new loss function based on the triangular representation of the standard I(1) cointegrating system leads to the superior ranking of the system forecasts over their univariate competitors, see expression (23).

4 Long-run forecasting in polynomially cointegrating systems.

Next, we examine the forecasting performance of the model that imposes the polynomially cointegrating restrictions and the model that totally ignores these when forecasting the original I(2) variables. Similarly to the last section we investigate the long-run behavior of the loss function based on the traditional trace MSFE criterion when comparing the forecasting performance of the model that imposes the polynomially cointegrating restrictions with that of the univariate model. We will show that also in this case,

[6] The problem is that the parameters θ_1 and σ_u^2 are functions of both the parameters of the multicointegrated model $(\lambda, \alpha, \sigma_1^2, \sigma_2^2)$ as well as the parameter θ_2, which is a solution to the fourth-order polynomial (9) derived from the MA(2) process characterizing the univariate representations, see equations given in (10).

imposing polynomially cointegrating restrictions does not improve over the long-run forecasting performance of the simple correctly specified univariate models. This well accords with the established results of Christoffersen and Diebold (1998) for the cointegrating systems as well as in the section above for the multicointegrating systems.

In order to combat this fact, we suggest a new loss function based on the triangular representation of the polynomially cointegrating variables given in (6). Opposite to the conventional trace MSFE criterion, the new loss function explicitly recognizes the important distinction between the system- and univariate forecasts. The system forecasts obey the polynomially cointegrating relations exactly in the limit, whereas this is not true in the case with the forecasts from the univariate representations.

4.1 Forecasts of I(2) variables from polynomially cointegrating system.

Using the MA-representation (7) presented in Section 2.2 we can write the future values of the I(2) variables

$$\begin{aligned} X_{t+h} &= X_t + h\Delta X_t + \sum_{q=1}^{h}\sum_{i=1}^{q} e_{1t+i} \\ \Delta X_{t+h} &= \Delta X_t + \sum_{i=1}^{h} e_{1t+i} \\ Y_{t+h} &= \lambda\left(X_t + h\Delta X_t\right) + \alpha\Delta X_t + \lambda\sum_{q=1}^{h}\sum_{i=1}^{q} e_{1t+i} + \alpha\sum_{i=1}^{h} e_{1t+i} + e_{2t+h} \end{aligned}$$

with the corresponding forecasts

$$\widehat{X}_{t+h} = X_t + h\Delta X_t \tag{25}$$

$$\Delta\widehat{X}_{t+h} = \Delta X_t \tag{26}$$

$$\widehat{Y}_{t+h} = \lambda\left(X_t + h\Delta X_t\right) + \alpha\Delta X_t. \tag{27}$$

The forecast errors read

$$\begin{aligned} \widehat{\varepsilon}_{X,t+h} &= \sum_{q=1}^{h}\sum_{i=1}^{q} e_{1t+i} = \sum_{i=1}^{h}\left(h+1-i\right) e_{1t+i} \\ \widehat{\varepsilon}_{\Delta X,t+h} &= \sum_{i=1}^{h} e_{1t+i} \\ \widehat{\varepsilon}_{Y,t+h} &= \sum_{i=1}^{h}\left[\lambda\left(h+1-i\right)+\alpha\right] e_{1t+i} + e_{2t+h} \end{aligned}$$

and follow the same stochastic process as the original system, that is

$$\begin{bmatrix} \Delta^2\widehat{\varepsilon}_{X,t+h} \\ \Delta^2\widehat{\varepsilon}_{Y,t+h} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda + \alpha\left(1-L\right) & \left(1-L\right)^2 \end{bmatrix} \begin{bmatrix} e_{1t+h} \\ e_{2t+h} \end{bmatrix}.$$

The corresponding forecast error variances for the levels of I(2) variables are of the order $O\left(h^3\right)$ as seen below:

$$
\begin{aligned}
Var\left(\widehat{\varepsilon}_{X,t+h}\right) &= Var\left(\sum_{q=1}^{h}\sum_{i=1}^{q} e_{1t+i}\right) = \frac{h\left(h+1\right)\left(2h+1\right)}{6}\sigma_1^2 \sim O\left(h^3\right) \\
Var\left(\widehat{\varepsilon}_{\Delta X,t+h}\right) &= \sum_{i=1}^{h} Var\left(e_{1t+i}\right) = h\sigma_1^2 \\
Var\left(\widehat{\varepsilon}_{Y,t+h}\right) &= \frac{h\left(h+1\right)\left(2h+1\right)}{6}\lambda^2\sigma_1^2 + 2\alpha\lambda\frac{h\left(h+1\right)}{2}\sigma_1^2 + h\alpha^2\sigma_1^2 + \sigma_2^2 \sim O\left(h^3\right)
\end{aligned}
$$

The variance of the polynomially cointegrating combination of the forecast errors reads

$$
Var\left(\widehat{\varepsilon}_{Y,t+h} - \lambda\widehat{\varepsilon}_{X,t+h} - \alpha\widehat{\varepsilon}_{\Delta X,t+h}\right) = \sigma_2^2. \tag{28}
$$

This is finite, and for our simple model it is constant for all forecast horizons $h > 0$ as there is no short-run dynamics. The finding of the finite variance of the polynomially cointegrating combination of the forecast errors is similar to that of Christoffersen and Diebold (1998), and Engle and Yoo (1987) for I(1) systems. This is due to the fact that the forecast errors follow the same stochastic process as the forecasted time series. As a consequence, the forecasts are integrated of the same order and share the polynomially cointegrating properties of the system dynamics as well.

In the model the forecasts satisfy exactly the polynomially cointegrating relation at all horizons, not just in the limit. This can be shown using the expressions $(25), (26)$ and (27) :

$$
\widehat{Y}_{t+h} - \lambda\widehat{X}_{t+h} - \alpha\Delta\widehat{X}_{t+h} = \lambda\left(X_t + h\Delta X_t\right) + \alpha\Delta X_t - \lambda\left(X_t + h\Delta X_t\right) - \alpha\Delta X_t = 0, \quad \text{for } h > 0
$$

4.2 Forecasts from the implied univariate representations for I(2) variables.

Next we derive the forecast expressions from the implied univariate representations. The future values of the process X_t are the same as based on the triangular system (4)

$$
\begin{aligned}
X_{t+h} &= X_t + h\Delta X_t + \sum_{q=1}^{h}\sum_{i=1}^{q} e_{1t+i} \\
\Delta X_{t+h} &= \Delta X_t + \sum_{i=1}^{h} e_{1t+i}
\end{aligned}
$$

and for the process Y_{t+h} we have

$$
Y_{t+h} = Y_t + h\Delta Y_t + h\left(u_{t+1} + \theta_1 u_t + \theta_2 u_{t-1}\right) + \left(h-1\right)\left(u_{t+2} + \theta_1 u_{t+1} + \theta_2 u_t\right) + \sum_{q=1}^{h-2}\sum_{i=1}^{q} z_{t+i+2}
$$

The corresponding forecasts for I(2) variables are calculated as follows. For X_t it is the same as from the polynomially cointegrating model

$$\begin{aligned}\widetilde{X}_{t+h} &= \widehat{X}_{t+h} = X_t + h\Delta X_t \\ \Delta\widetilde{X}_{t+h} &= \Delta\widehat{X}_{t+h} = \Delta X_t\end{aligned}$$

The univariate forecast for Y_{t+h} reads

$$\widetilde{Y}_{t+h} = \begin{cases} Y_t + \Delta Y_t + \theta_1 u_t + \theta_2 u_{t-1} & \text{for } h = 1 \\ Y_t + h\Delta Y_t + h(\theta_1 + \theta_2) u_t + h\theta_2 u_{t-1} - \theta_2 u_t & \text{for } h > 1 \end{cases}.$$

Notice that in this case the long-run forecasts from the implied univariate representations do not maintain the polynomially cointegrating relations in the long-run exactly but do so on average as opposed to their system counterparts that maintain the polynomially cointegrating relations exactly in the long-run. To see this, we have

$$\begin{aligned}\widetilde{Y}_{t+h} - \lambda\widetilde{X}_{t+h} - \alpha\Delta\widetilde{X}_{t+h} &= Y_t + h\Delta Y_t + h(\theta_1 + \theta_2) u_t + h\theta_2 u_{t-1} - \theta_2 u_t - \lambda X_t - \lambda h\Delta X_t - \alpha\Delta X_t \\ &= [Y_t - \lambda X_t - \alpha\Delta X_t] + h[\Delta Y_t - \lambda\Delta X_t] + h(\theta_1 + \theta_2) u_t + h\theta_2 u_{t-1} - \theta_2 u_t.\end{aligned}$$

The corresponding forecast errors are given by

$$\begin{aligned}\widetilde{\varepsilon}_{X,t+h} &= \widehat{\varepsilon}_{X,t+h} = \sum_{q=1}^{h}\sum_{i=1}^{q} e_{1t+i} = \sum_{i=1}^{h}(h+1-i)\, e_{1t+i} \\ \widetilde{\varepsilon}_{\Delta X,t+h} &= \widehat{\varepsilon}_{\Delta X,t+h} = \sum_{i=1}^{h} e_{1t+i}\end{aligned}$$

with the variances

$$Var(\widetilde{\varepsilon}_{X,t+h}) = Var(\widehat{\varepsilon}_{X,t+h}) = \frac{h(h+1)(2h+1)}{6}\sigma_1^2 \sim O\left(h^3\right) \tag{29}$$

$$Var(\widetilde{\varepsilon}_{\Delta X,t+h}) = Var(\widehat{\varepsilon}_{\Delta X,t+h}) = h\sigma_1^2 \sim O(h). \tag{30}$$

The forecast errors for Y_{t+h} read

$$\begin{aligned}\widetilde{\varepsilon}_{Y,t+1} &= u_{t+1} \\ \widetilde{\varepsilon}_{Y,t+h} &= \sum_{i=1}^{h-2}\{(1+\theta_1+\theta_2)(h-i+1) + (1+\theta_1) + 1\} u_{t+i} + ((1+\theta_1)+1) u_{t+h-1} + u_{t+h}\end{aligned}$$

with the corresponding forecast error variances

$$\begin{aligned}Var(\widetilde{\varepsilon}_{Y,t+1}) &= \sigma_u^2 \\ Var(\widetilde{\varepsilon}_{Y,t+h}) &= (1+\theta_1+\theta_2)^2 \frac{(h-2)(h-2+1)(2(h-2)+1)}{6}\sigma_u^2 \\ &\quad +2((1+\theta_1)+1)(1+\theta_1+\theta_2)\frac{(h-2)(h-2+1)}{2}\sigma_u^2 - \\ &\quad +((1+\theta_1)+1)^2(h-1)\sigma_u^2 + \sigma_u^2 \sim O\left(h^3\right).\end{aligned} \tag{31}$$

Next, we calculate the variance of the polynomially cointegrating combination of the forecast errors from the univariate representation. The straightforward but tedious algebra relegated to the appendix yields the following relation

$$Var\left(\widetilde{\varepsilon}_{Y,t+h}-\lambda\widetilde{\varepsilon}_{X,t+h}-\alpha\widetilde{\varepsilon}_{\Delta X,t+h}\right)=\left[Var\left(\widetilde{\varepsilon}_{Y,t+h}\right)-Var\left(\widehat{\varepsilon}_{Y,t+h}\right)\right]+\sigma_2^2. \tag{32}$$

Moreover, as shown in the appendix the leading term for the h^3 eventually cancels out such that the variance of the polynomially cointegrating combination of the forecast errors from the univariate representation is growing of order $O\left(h^2\right)$. That is, it grows at a lower order than the variance of the forecast errors for the levels of I(2) variables.

Thus, the fact, that the system long-run forecasts preserve the polynomially cointegrating relations exactly whereas the univariate long-run forecasts do so only on average, allows us to construct a new loss function in the spirit of Christoffersen and Diebold (1998), which takes into account this important distinction between the system- and univariate forecasts.

4.3 Comparison of forecast accuracy for I(2) variables.

First, we show that the ratio of the usual trace MSFE for univariate and system forecast errors tend to unity as the forecast horizon increases. Using expressions (29), (30), and (31) we can derive the following result

$$\frac{trace(Var(\widetilde{\varepsilon}_{t+h}))}{trace(Var(\widehat{\varepsilon}_{t+h}))}=\frac{Var\left(\widetilde{\varepsilon}_{X,t+h}\right)+Var\left(\widetilde{\varepsilon}_{Y,t+h}\right)}{Var\left(\widehat{\varepsilon}_{X,t+h}\right)+Var\left(\widehat{\varepsilon}_{Y,t+h}\right)}=\frac{O\left(h^3\right)}{O\left(h^3\right)}\rightarrow 1 \tag{33}$$

since the coefficients to the leading terms are identical. Observe, that this result is related to the expression (32), where these equivalent coefficients to the leading terms resulted in cancellation of those leading terms, and thus reducing the growth order of the variance of the polynomially cointegrating combination of the univariate forecast errors from $O\left(h^3\right)$ to $O\left(h^2\right)$.

In contrast, the ratio of trace MSFE for the triangular representation of the polynomially cointegrating system does not tend to unity but diverges to infinity as the forecast horizon increases. For the system forecasts we have

$$\begin{aligned} trace\ \widehat{\mathrm{MSFE}}_{tri} &= E\left\{\begin{pmatrix}\widehat{\varepsilon}_{Y,t+h}-\lambda\widehat{\varepsilon}_{X,t+h}-\alpha\widehat{\varepsilon}_{\Delta X,t+h}\\ (1-L)^2\,\widehat{\varepsilon}_{X,t+h}\end{pmatrix}'\begin{pmatrix}\widehat{\varepsilon}_{Y,t+h}-\lambda\widehat{\varepsilon}_{X,t+h}-\alpha\widehat{\varepsilon}_{\Delta X,t+h}\\ (1-L)^2\,\widehat{\varepsilon}_{X,t+h}\end{pmatrix}\right\}= \\ &= E\left(\widehat{\varepsilon}_{Y,t+h}-\lambda\widehat{\varepsilon}_{X,t+h}-\alpha\widehat{\varepsilon}_{\Delta X,t+h}\right)^2+E\left((1-L)^2\,\widehat{\varepsilon}_{X,t+h}\right)^2=\sigma_2^2+\sigma_1^2=O\left(1\right)\end{aligned}$$

and for the univariate forecasts we have that

$$\begin{aligned} trace\ \widetilde{\text{MSFE}}_{tri} &= E\left\{ \begin{pmatrix} \widetilde{\varepsilon}_{Y,t+h} - \lambda\widetilde{\varepsilon}_{X,t+h} - \alpha\widetilde{\varepsilon}_{\Delta X,t+h} \\ (1-L)^2 \widetilde{\varepsilon}_{X,t+h} \end{pmatrix}' \begin{pmatrix} \widetilde{\varepsilon}_{Y,t+h} - \lambda\widetilde{\varepsilon}_{X,t+h} - \alpha\widetilde{\varepsilon}_{\Delta X,t+h} \\ (1-L)^2 \widetilde{\varepsilon}_{X,t+h} \end{pmatrix} \right\} = \\ &= E\left(\widetilde{\varepsilon}_{Y,t+h} - \lambda\widetilde{\varepsilon}_{X,t+h} - \alpha\widetilde{\varepsilon}_{\Delta X,t+h}\right)^2 + E\left((1-L)^2 \widetilde{\varepsilon}_{X,t+h}\right)^2 = O\left(h^2\right) + O\left(1\right) \end{aligned}$$

Hence, the ratio of trace $\widehat{\text{MSFE}}_{tri}$ and trace $\widetilde{\text{MSFE}}_{tri}$ is

$$\frac{trace\ \widetilde{\text{MSFE}}_{tri}}{trace\ \widehat{\text{MSFE}}_{tri}} = \frac{O\left(h^2\right)}{O\left(1\right)} \to \infty \quad \text{as} \quad h \to \infty. \tag{34}$$

This means that we would prefer the model with polynomially cointegrating restrictions using this criterion. In fact, there are high (increasing) gains to be achieved in using the new loss function over the traditional one.

Using equation (32) we have the following result

$$\frac{trace\ \widetilde{\text{MSFE}}_{tri}}{trace\ \widehat{\text{MSFE}}_{tri}} = \frac{\left[Var\left(\widetilde{\varepsilon}_{Y,t+h}\right) - Var\left(\widehat{\varepsilon}_{Y,t+h}\right)\right] + \sigma_2^2 + \sigma_1^2}{\sigma_2^2 + \sigma_1^2} = 1 + \frac{\left[Var\left(\widetilde{\varepsilon}_{Y,t+h}\right) - Var\left(\widehat{\varepsilon}_{Y,t+h}\right)\right]}{\sigma_2^2 + \sigma_1^2} > 1.$$

Intuitively, this inequality holds as the forecasts that utilize all the information in the system (system forecasts) will produce a smaller forecast error variance than the ones that are based on the partial information (univariate forecasts).

5 Example.

We illustrate the findings of the previous sections using the model (4) with the following values of the parameters $\lambda = 2, \alpha = 1, \sigma_1^2 = \sigma_2^2 = 1$. Such parameter combination leads to the following values of MA(2) process of $\Delta^2 Y_t = \Delta y_t : \theta_1 = -0.5155$, $\theta_2 = 0.0795$, and $\sigma_u^2 = 12.578$. Figures 1 and 2 below are plotted using these true coefficient parameters. Figure 1 displays the ratios (22), (24) and (23). Similarly, Figure 2 corresponds to the results given in (33) and (34). As seen, the analytical findings are nicely verified in the graphs.

6 Conclusions.

In the present paper we have investigated the issue of long-run forecasting in the systems with multi– and polynomially cointegrating restrictions. We have showed that the results of Christoffersen and Diebold (1998) derived for the standard I(1) cointegrating systems generally hold also for the present systems of focus. That is, on the basis of the loss function based on the traditional trace MSFE criterion, imposing

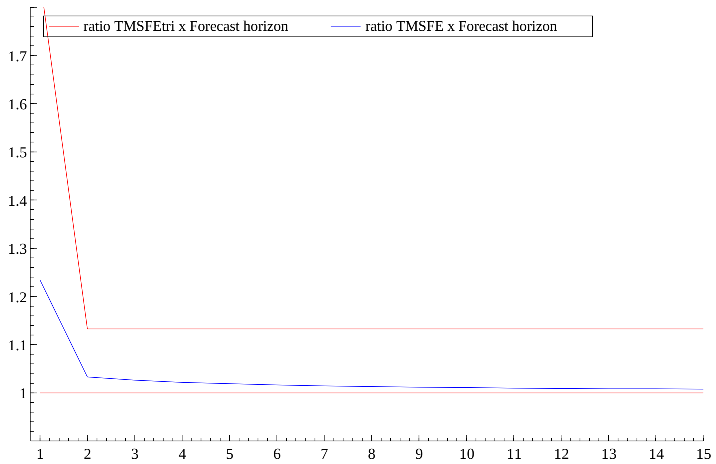

Figure 1: Trace MSFE ratio and Trace MSFE$_{tri}$ ratio of univariate versus system forecasts of multicointegrating I(1) variables.

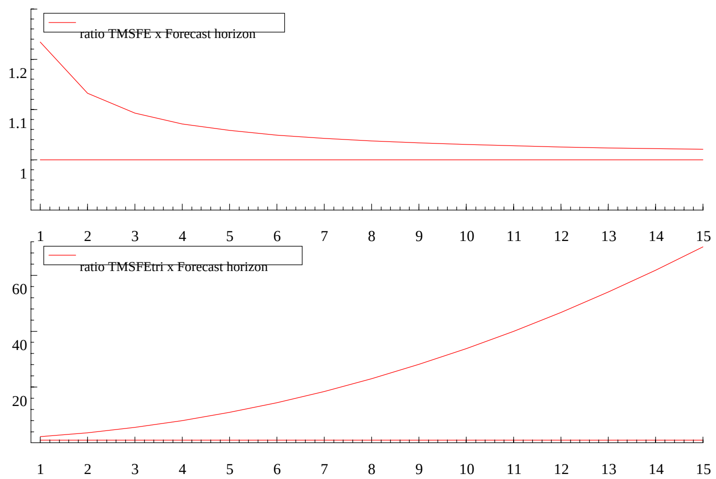

Figure 2: Trace MSFE ratio and Trace MSFE$_{tri}$ ratio of univariate versus system forecasts of polynomially cointegrating I(2) variables.

the relevant restrictions does not lead to the improved long-run forecast performance when compared to the forecasting from the simple univariate models.

However, the clear distinction between the system- and univariate forecasts can be achieved if one employs the loss function based on the triangular representation of the cointegrating and polynomially cointegrating models. In this case, the measurable gains come from the fact that this particular loss function explicitly acknowledges the important distinction between the system- and univariate forecasts. The intrinsic feature of the system forecasts is that they maintain the (polynomially-) cointegrating restrictions in the limit exactly, whereas this is not so in case of univariate long-run forecasts. Hence, the paper highlights the importance of carefully selecting loss functions when evaluating forecasts from cointegrating systems.

In this paper we used a simple multi- and polynomially cointegrating models in order to establish the results. Naturally, it is of interest to derive the corresponding results for the general models that obey multi- and polynomially cointegrating restrictions. Also, the consequences of introducing deterministic components is of importance as are estimation issues. These extensions will follow in future work.

References

Banerjee, A., L. Cockerell, and B. Russell (2001): "An I(2) Analysis of Inflation and the Markup," *Journal of Applied Econometrics*, 16(3), 221–40.

Campbell, J. Y., and R. J. Shiller (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1052–1088.

Christoffersen, P. F., and F. X. Diebold (1998): "Cointegration and Long-Run Forecasting," *Journal of Business and Economic Statistics*, 16(4), 450–458.

Clements, M. P., and D. F. Hendry (1995): "Forecasting in Cointegrating Systems," *Journal of Applied Econometrics*, 10(2), 127–146.

Engle, R. F., and B. S. Yoo (1987): "Forecasting and Testing in Co-Integrated Systems," *Journal of Econometrics*, 35, 143–159.

Engsted, T., J. Gonzalo, and N. Haldrup (1997): "Testing for Multicointegration," *Economic Letters*, 56, 259–266.

Engsted, T., and N. Haldrup (1999): "Multicointegration in Stock-Flow Models," *Oxford Bulletin of Economics and Statistics*, 61, 237–254.

ENGSTED, T., AND S. JOHANSEN (1999): "Granger's Representation Theorem and Multicointegration," in *Festschrift Cointegration, Causality and Forecasting, Festschrift in Honour of Clive Granger*, ed. by R. Engle, and H.White, Oxford. Oxford University Press.

GRANGER, C. W. J., AND T. H. LEE (1989): "Investigation of Production, Sales and Inventory Relations Using Multicointegration and Non-Symmetric Error Correction Models," *Journal of Applied Econometrics*, 4, S145–S159.

GRANGER, C. W. J., AND T. H. LEE (1991): "Multicointegration," in *Long-Run Economic Relationships. Reading in Cointegration*, ed. by R. F. Engle, and C. W. J. Granger, Advanced Texts in Econometrics, Oxford. Oxford University Press.

HALDRUP, N. (1998): "A Review of the Econometric Analysis of I(2) Variables," *Journal of Economic Surveys*, 12(5), 595–650.

HENDRY, D. F., AND T. VON UNGERN-STERNBERG (1981): "Liquidity and Inflation Effects on Consumers' Expenditure," in *Essay in the Theory and Measurement of Consumers' Behaviour*, ed. by A. S. Deaton, Cambridge. Cambridge University Press.

LEACHMAN, L. L. (1996): "New Evidence on the Ricardian Equivalence Theorem: A Multicointegration Approach," *Applied Economics*, 28(6), 695–704.

LEACHMAN, L. L., AND B. B. FRANCIS (2000): "Multicointegration Analysis of the Sustainability of Foreign Debt," *Journal of Macroeconomics*, 22(2), 207–27.

LEE, T. H. (1992): "Stock-Flow Relationships in US Housing Construction," *Oxford Bulletin of Economics and Statistics*, 54, 419–430.

PHILLIPS, P. C. B. (1991): "Optimal Inference in Cointegrating Systems," *Econometrica*, 59, 283–306.

RAHBEK, A., H. C. KONGSTED, AND C. JØRGENSEN (1999): "Trend-Stationarity in the I(2) Cointegration Model," *Journal of Econometrics*, 90, 265–289.

SILIVERSTOVS, B. (2001): "Multicointegration in US Consumption Data," Aarhus University, Department of Economics, Working paper 2001-6.

7 Appendix.

7.1 Derivation of the implied univariate representation for Δy_t and $\Delta^2 Y_t$.

$$\begin{aligned} z_t &= [\lambda + (1-L)\alpha] e_{1t} + (1-L)^2 e_{2t} \\ z_t &= \lambda e_{1t} + \alpha e_{1t} - \alpha e_{1t-1} + e_{2t} - 2e_{2t-1} + e_{2t-2}. \\ z_t &= u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} \end{aligned}$$

The autocovariance structure for z_t reads

$$\begin{aligned} \gamma_z(0) &= \left[(\lambda+\alpha)^2 + \alpha^2\right]\sigma_1^2 + 6\sigma_2^2 \\ \gamma_z(1) &= -\alpha(\lambda+\alpha)\sigma_1^2 - 4\sigma_2^2 \\ \gamma_z(2) &= \sigma_2^2 \\ \gamma_z(\tau) &= 0, \quad |\tau| \geq 3. \end{aligned}$$

This is a MA(2) process with the non-zero first and second autocorrelations. The first autocorrelation coefficient is

$$\begin{aligned} \rho_z(1) &= \frac{-\alpha(\lambda+\alpha)\sigma_1^2 - 4\sigma_2^2}{\left[(\lambda+\alpha)^2+\alpha^2\right]\sigma_1^2 + 6\sigma_2^2} = \frac{-\alpha(\lambda+\alpha)q - 4}{\left[(\lambda+\alpha)^2+\alpha^2\right]q + 6} \\ \rho_z(2) &= \frac{\sigma_2^2}{\left[(\lambda+\alpha)^2+\alpha^2\right]\sigma_1^2 + 6\sigma_2^2} = \frac{1}{\left[(\lambda+\alpha)^2+\alpha^2\right]q + 6}, \end{aligned}$$

where

$$q = \frac{\sigma_1^2}{\sigma_2^2}$$

is the signal-to-noise ratio.

From this we can try to infer values for the parameters θ_1 and θ_2. By denoting

$$A = [-\alpha(\lambda+\alpha)q - 4] \qquad B = \left[(\lambda+\alpha)^2 + \alpha^2\right]q + 6$$

and after some algebra we have that

$$\theta_1 = \frac{\theta_2}{(1+\theta_2)}A$$

and θ_2 is one of the root of the fourth-order polynomial

$$\theta_2^4 + (2-B)\theta_2^3 + \left(A^2 - 2B + 2\right)\theta_2^2 + (2-B)\theta_2 + 1 = 0.$$

Observe that the coefficient values θ_1 and θ_2 should satisfy the invertibility conditions for the MA(2) process z_t. The variance σ_u^2 is found from the following expression

$$\sigma_u^2 = \frac{\left[(\lambda+\alpha)^2+\alpha^2\right]\sigma_1^2+6\sigma_2^2}{\left(1+\theta_1^2+\theta_2^2\right)} \quad \text{or} \quad \sigma_u^2 = \frac{\sigma_2^2}{\theta_2}.$$

Furthermore, the following interesting relation holds

$$\frac{(1+\theta_1+\theta_2)^2}{\left(1+\theta_1^2+\theta_2^2\right)} = \frac{\lambda^2\sigma_1^2}{\left[(\lambda+\alpha)^2+\alpha^2\right]\sigma_1^2+6\sigma_2^2},$$

which further leads to

$$\lambda^2\sigma_1^2 = (1+\theta_1+\theta_2)^2\,\sigma_u^2.$$

7.2 Variance of the polynomially cointegrating combination of univariate forecast errors.

Here, we calculate the variance of the polynomially cointegrating combination of the forecast errors from the univariate representation:

$$\begin{aligned}
&Var\left(\widetilde{\varepsilon}_{Y,t+h}-\lambda\widetilde{\varepsilon}_{X,t+h}-\alpha\widetilde{\varepsilon}_{\Delta X,t+h}\right)= \\
&=Var\left(\widetilde{\varepsilon}_{Y,t+h}-\lambda\widetilde{\varepsilon}_{X,t+h}\right)+\alpha^2 Var\left(\widetilde{\varepsilon}_{\Delta X,t+h}\right)-2\alpha Cov\left(\widetilde{\varepsilon}_{Y,t+h}-\lambda\widetilde{\varepsilon}_{X,t+h},\widetilde{\varepsilon}_{\Delta X,t+h}\right)= \\
&=Var\left(\widetilde{\varepsilon}_{Y,t+h}\right)+\lambda^2 Var\left(\widetilde{\varepsilon}_{X,t+h}\right)-2\lambda Cov\left(\widetilde{\varepsilon}_{Y,t+h},\widetilde{\varepsilon}_{X,t+h}\right)+\alpha^2 Var\left(\widetilde{\varepsilon}_{\Delta X,t+h}\right)- \\
&-2\alpha Cov\left(\widetilde{\varepsilon}_{Y,t+h}\widetilde{\varepsilon}_{\Delta X,t+h}\right)+2\alpha\lambda Cov\left(\widetilde{\varepsilon}_{X,t+h},\widetilde{\varepsilon}_{\Delta X,t+h}\right).
\end{aligned}$$

Thus, in order to calculate the variance of the polynomially cointegrating combination of the forecast errors we need to derive the following expressions:

$$\begin{aligned}
Var\left(\widetilde{\varepsilon}_{Y,t+h}\right) &= (1+\theta_1+\theta_2)^2\frac{(h-2)(h-2+1)(2(h-2)+1)}{6}\sigma_u^2 \\
&\quad +2\left((1+\theta_1)+1\right)(1+\theta_1+\theta_2)\frac{(h-2)(h-2+1)}{2}\sigma_u^2- \\
&\quad +\left((1+\theta_1)+1\right)^2(h-1)\sigma_u^2+\sigma_u^2 \\
Var\left(\widetilde{\varepsilon}_{X,t+h}\right) &= Var\left(\sum_{q=1}^{h}\sum_{i=1}^{q}e_{1t+i}\right)=\left(h^2+(h-1)^2+..+1\right)\sigma_1^2=\frac{h(h+1)(2h+1)}{6}\sigma_1^2 \\
Var\left(\widetilde{\varepsilon}_{\Delta X,t+h}\right) &= h\sigma_1^2
\end{aligned}$$

$$\begin{aligned}
Cov\left(\widetilde{\varepsilon}_{Y,t+h},\widetilde{\varepsilon}_{X,t+h}\right) &= \lambda\frac{h(h+1)(2h+1)}{6}\sigma_1^2+\alpha\frac{h(h+1)}{2}\sigma_1^2 \\
Cov\left(\widetilde{\varepsilon}_{Y,t+h},\widetilde{\varepsilon}_{\Delta X,t+h}\right) &= \lambda\frac{h(h+1)}{2}\sigma_1^2+\alpha h\sigma_1^2 \\
Cov\left(\widetilde{\varepsilon}_{X,t+h},\widetilde{\varepsilon}_{\Delta X,t+h}\right) &= Cov\left(\sum_{q=1}^{h}\sum_{i=1}^{q}e_{1t+i},\sum_{i=1}^{h}e_{1t+i}\right)=\frac{h(h+1)}{2}\sigma_1^2
\end{aligned}$$

These expressions lead to the following result:

$$\begin{aligned}
&Var\left(\widetilde{\varepsilon}_{Y,t+h}-\lambda\widetilde{\varepsilon}_{X,t+h}-\alpha\widetilde{\varepsilon}_{\Delta X,t+h}\right)=\\
&=Var\left(\widetilde{\varepsilon}_{Y,t+h}\right)-\lambda^{2}\tfrac{h(h+1)(2h+1)}{6}\sigma_1^2-2\alpha\lambda\tfrac{h(h+1)}{2}\sigma_1^2-\alpha^2h\sigma_1^2=\\
&=Var\left(\widetilde{\varepsilon}_{Y,t+h}\right)-\left[Var\left(\widehat{\varepsilon}_{Y,t+h}\right)-\sigma_2^2\right]=\left[Var\left(\widetilde{\varepsilon}_{Y,t+h}\right)-Var\left(\widehat{\varepsilon}_{Y,t+h}\right)\right]+\sigma_2^2
\end{aligned}$$

and some further simplification leads to

$$\begin{aligned}
&Var\left(\widetilde{\varepsilon}_{Y,t+h}-\lambda\widetilde{\varepsilon}_{X,t+h}-\alpha\widetilde{\varepsilon}_{\Delta X,t+h}\right)=\\
&=-\lambda^2 2h\left(h+1\right)\sigma_1^2+\left(\left(1+\theta_1\right)+1\right)\left(1+\theta_1+\theta_2\right)h\left(h+1\right)\sigma_u^2\\
&+4\lambda^2h\sigma_1^2-4\left(\left(1+\theta_1\right)+1\right)\left(1+\theta_1+\theta_2\right)h\sigma_u^2+\left(\left(1+\theta_1\right)+1\right)^2h\sigma_u^2\\
&-\lambda^2\sigma_1^2+2\left(\left(1+\theta_1\right)+1\right)\left(1+\theta_1+\theta_2\right)\sigma_u^2-\left(\left(1+\theta_1\right)+1\right)^2\sigma_u^2+\sigma_u^2-\alpha\lambda h(h+1)\sigma_1^2-\alpha^2h\sigma_1^2.
\end{aligned}$$

As seen the variance of the multicointegrating combination of the forecasts errors from the univariate models is of growth order $O\left(h^2\right)$. The last expression also reads as follows

$$\begin{aligned}
&Var\left(\widetilde{\varepsilon}_{Y,t+h}-\lambda\widetilde{\varepsilon}_{X,t+h}-\alpha\widetilde{\varepsilon}_{\Delta X,t+h}\right)=\\
&=-\lambda^2 2h\left(h+1\right)\sigma_1^2+\left(\left(1+\theta_1\right)+1\right)\left(1+\theta_1+\theta_2\right)h\left(h+1\right)\sigma_u^2+\left[2\left(1+\theta_1+\theta_2\right)-\left(\left(1+\theta_1\right)+1\right)\right]h\sigma_u^2-\\
&-\left[\left(1+\theta_1+\theta_2\right)-\left(\left(1+\theta_1\right)+1\right)\right]\sigma_u^2+\sigma_u^2-\alpha\lambda h(h+1)\sigma_1^2-\alpha^2h\sigma_1^2\\
&=-\lambda^2 2h\left(h+1\right)\sigma_1^2+\left(\left(1+\theta_1\right)+1\right)\left(1+\theta_1+\theta_2\right)h\left(h+1\right)\sigma_u^2+\\
&+\left[\theta_1+2\theta_2\right]h\sigma_u^2+\left[1-\theta_2\right]\sigma_u^2+\sigma_u^2-\alpha\lambda h(h+1)\sigma_1^2-\alpha^2h\sigma_1^2
\end{aligned}$$