ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Forecasting Accuracy and Estimation Uncertainty Using VAR Models with Short- and Long-Term Economic Restrictions: A Monte-Carlo Study

Osmani Teixeira de Carvalho Guillén,
João Victor Issler and George Athanasopoulos

May 2005

Working Paper 15/05

Forecasting Accuracy and Estimation Uncertainty using VAR Models with Short- and Long-Term Economic Restrictions: A Monte-Carlo Study*

Osmani Teixeira de Carvalho Guillén
Banco Central do Brasil
Av. Presidente Vargas, 730 - Centro
Rio de Janeiro, RJ 20071-001
Brazil

João Victor Issler†
Graduate School of Economics – EPGE
Getulio Vargas Foundation
Praia de Botafogo 190 s. 1111
Rio de Janeiro, RJ 22253-900
Brazil
jissler@fgv.br

George Athanasopoulos
Department of Econometrics and Business Statistics
Monash University, Clayton, Victoria 3800
Australia

March, 2005.

Abstract

Using vector autoregressive (VAR) models and Monte-Carlo simulation methods we investigate the potential gains for forecasting accuracy and estimation uncertainty of two commonly used restrictions arising from economic relationships. The first reduces parameter space by imposing long-term restrictions on the behavior of economic variables as discussed by the literature on cointegration, and the second reduces parameter space by imposing short-term restrictions as discussed by the literature on

*Acknowledgments: Parts os this paper were writen while Osmani T. Guillén and João Victor Issler were visiting Monash University, which hospitality is gratefully acknowledged. João Victor Issler and George Athanasopulos also acknowledge the hospitality of the Australian National University, where parts of this paper were writen. We gratefully acknowledge comments and suggestions given by Alain Hecq, Luiz Renato Lima, and of the participants of the conferences *Common Features in London* and *Encontro Brasileiro de Econometria.* Special thanks are due to Farshid Vahid for some of the ideas in this paper and his encouragement and support to Osmani T. Guillén during his visit to Monash University and to George Athanasopoulos. The usual disclaimers apply. João Victor Issler and Osmani T. Guillén acknowledge, respectively, the support of CNPq-Brazil, PRONEX, and CAPES fellowship BEX0934/02-0.

†Corresponding author.

serial-correlation common features (SCCF). Our simulations cover three important issues on model building, estimation, and forecasting. First, we examine the performance of standard and modified information criteria in choosing lag length for cointegrated VARs with SCCF restrictions. Second, we provide a comparison of forecasting accuracy of fitted VARs when only cointegration restrictions are imposed and when cointegration and SCCF restrictions are jointly imposed. Third, we propose a new estimation algorithm where short- and long-term restrictions interact to estimate the cointegrating and the cofeature spaces respectively.

We have three basic results. First, ignoring SCCF restrictions has a high cost in terms of model selection, because standard information criteria chooses too frequently inconsistent models, with too small a lag length. Criteria selecting lag and rank simultaneously have a superior performance in this case. Second, this translates into a superior forecasting performance of the restricted VECM over the VECM, with important improvements in forecasting accuracy – reaching more than 100% in extreme cases. Third, the new algorithm proposed here fares very well in terms of parameter estimation, even when we consider the estimation of long-term parameters, opening up the discussion of joint estimation of short- and long-term parameters in VAR models.

Keywords: Reduced rank models, model selection criteria, forecasting accuracy.

JEL Classification: C32, C53.

1 Introduction

One of the objectives of time-series econometrics is to identify relationships between economic variables and then use them for estimation and forecasting. In theory, an identified relationship should improve the accuracy of forecasts through the reduction of estimation uncertainty. However, in practice, the ultimate gains of this procedure may be small or nonexistent, because of the risk of model misspecification at several stages of model selection; see the discussion on cointegration testing in Lin and Tsay(1996), and the discussion on VAR-order and rank selection in Vahid and Issler(2002). A reason for model misspecification in the context of VAR models is given in Johansen(1991) and Gonzalo(1994), who point out that VAR-order selection may affect proper inference on cointegrating vectors and rank.

In this paper, we take vector autoregressive (VAR) models, which have become the "working horses" for macroeconometric studies, and investigate the potential gains for forecasting accuracy and estimation uncertainty of two commonly used restrictions arising from economic relationships. The first reduces parameter space by imposing long-term restrictions on the behavior of economic variables as discussed by the literature on cointegration after Granger(1981), Engle and Granger(1987), and Johansen(1988, 1991). The second reduces parameter space by imposing short-term restrictions as discussed by the literature on serial-correlation common features (SCCF) after Engle and Kozicki(1993), Vahid and Engle(1993, 1997) and Hecq, Palm and Urbain(2005).

There is a dichotomy between these two sets of rank restrictions, since cointegration restricts the low-frequency behavior of economic time-series and SCCF restricts their high-frequency behavior. One should expect that the gains of imposing cointegrating restrictions are realized for long-term forecasts of the series being modelled, whereas those of imposing SCCF restrictions are realized for their short-term forecasts. Indeed, this is a fundamental difference between them and it works toward the benefit of focusing on short-term restrictions. In the simulations of Engle and Yoo(1987), unconstrained VAR models produced better short-term forecasts than cointegrated VARs. Despite this shortcoming, the importance of cointegration for long-term forecasts was stressed by Lin and Tsay on a theoretical basis. However, their simulation and empirical results were not very encouraging, something Clements and Hendry(1995) concur with. Because forecasting uncertainty at long horizons can be large, time-series models are generally most useful for forecasting at short horizons. Hence, imposing short-term constraints are a way of improving the effectiveness of time-series models at horizons where they are most useful.

As far as we are aware of, this paper is the first to make a direct forecasting comparison between the consequences of imposing short- and long-term constraints on VAR models. Although there has been a considerable effort examining the importance of cointegration restrictions in VAR forecasting – see, among others, Engle and Yoo, Clements and Hendry, Lin and Tsay, Hoffman and Rasche(1996), Christoffersen and Diebold(1998), Diebold and Kilian(2001), and Silverstovs, Engsted and Haldrup(2004) – with the exception of Vahid and Issler there has been no thorough examination of the forecasting importance of common-cyclical features on the same scale. Even in the latter, data was assumed to be $I(0)$, therefore only SCCF restrictions were considered as a potential source for improving forecasting accuracy. There is an urge to make this direct comparison, since initial simulation and empirical results using cointegration restrictions have been discouraging while the opposite has happened when SCCF restrictions were considered.

As shown by Vahid and Issler, short-term SCCF restrictions impose linear constraints in forecasts at every forecasting horizon. If we apply this logic to a cointegrated VAR with SCCF restrictions, we conclude that it should outperform a cointegrated VAR in every finite horizon. Moreover, in the infinite horizon, their performance should be identical, since both impose the same long-term restrictions on the data. Despite that, there is always the risk of imposing false restrictions, which calls for conservative behavior: it is presumably better to use a possibly inefficient model instead of risking using an inconsistent model in the search for parsimony. Recent Monte-Carlo results in Vahid and Issler challenge this view with respect to VAR models with common cycles. They argue that the cost of ignoring common-cycle restrictions is more than just the efficiency loss. This happens because the usual practice in applied work of choosing lag length by information criteria will severely underparameterize in this case. For example, even for a relatively large sample size of 200 observations, the Akaike, Hanan-Quinn, and Schwarz criteria choose respectively a model with too small a lag length 55.7%, 95% and 99.9% of the time. For such misspecified models, there is little to learn from theory, except that all estimates are inconsistent.

Our simulations cover three important issues on model building, estimation, and forecasting. First, we examine the performance of standard information criteria ($IC(p)$) in choosing lag length p for cointegrated VARs with SCCF restrictions. The consequences of this performance for the estimation of long-term parameters is also investigated. We also compare the performance of $IC(p)$ with that of information criteria that chooses simultaneously lag length and the rank of the dynamic system r – $(IC(p, r))$. Second, we provide a comparison of forecasting accuracy of fitted VARs when only cointegration restrictions are imposed and when cointegration and SCCF restrictions are jointly imposed. For the sake of completeness, we also make comparisons with the forecasting performance of unrestricted VARs. These comparisons take into account the possibility of model misspecification in choosing the lag length of the VAR, the number of cointegrating vectors, and the number of coefeature vectors. Third, independently from Hecq(2005), we propose a new estimation algorithm where short- and long-term restrictions interact to estimate jointly the cointegrating and the cofeature spaces respectively. This algorithm follows closely the idea of weak-form reduced-rank structure suggested by Hecq, Palm and Urbain(2005). There, the reduced-rank structure of the lagged coefficient matrices in the cointegrated VAR is different from that of the adjustment coefficient matrix. The first pass of our algorithm only imposes weak-form SCCF restrictions, without imposing any restrictions on cointegrating rank. Based on first-pass restricted estimates, the long-run impact matrix is estimated without any rank constraints. The algorithm runs until there is convergence of short- and long-term coefficient estimates. At the end, it is possible to conduct inference on the cointegrating rank for the system. Here we inverted the usual order of estimation of VAR coefficients. The usual practice is to estimate cointegrating (rank) vectors first, and then, conditional on them, estimate the short-term dynamics of the system. Here, we estimate first the short-term dynamics, with no long-term constraints, only conducting cointegration inference and estimation at the end. We also provide a smaller simulation study examining the accuracy in estimating cointegrating vectors using this new algorithm, which allows a comparison with the method proposed by Johansen(1988, 1991).

The current study extends the work of Vahid and Issler in two dimensions. First, the unrestricted model being analyzed here is a cointegrated VAR, whereas in Vahid and Issler unit-root and cointegration restrictions were ignored, i.e., series were $I(0)$. Since cointegration is a common occurrence in macroeconomic models and data, we provide more relevant information about VAR models with SCCF restrictions than initial studies. Second, we propose a new estimation algorithm, where short- and long-term restrictions interact to estimate jointly the cointegrating and the cofeature spaces respectively. Because of its superior performance in small samples, we believe that it has the potential to form the basis of a future estimator for cointegrated VAR coefficients in the presence of SCCF restrictions.

The outline of the paper is as follows. Section 2 states the reduced-rank restrictions that common-cyclical fluctuations impose on the parameters of cointegrated VAR models, and discusses their consequences for forecasting. Section 3 describes in detail the new estimation algorithm proposed here for VAR models with short- and long-term restrictions. Section 4 describes our Monte-Carlo design; see also the discussion in the Appendix on DGP selection. Section 5 presents the simulation results and Section 6 concludes.

2 Theory and forecasting with restricted VAR models

2.1 Theory

In this section we present a brief discussion of VAR models with cointegration and SCCF restrictions. We focus on the autoregressive representation. For a complete discussion readers are referred to Engle and Granger(1987), Johansen(1991), Vahid and Engle(1993) and Hecq, Palm and Urbain(2005). We assume that y_t is a $n \times 1$ random vector generated by a Vector Autoregression (VAR) of order p:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t. \tag{1}$$

Cointegration implies that the matrix $I - \sum_{i=1}^{p} A_i$ must have less than full rank, which imposes cross-equation restrictions on the VAR, which can be written as a Vector Error-Correction model (VECM):

$$\begin{aligned}
\Delta y_t &= A_1^* \Delta y_{t-1} + \ \ldots \ + A_{p-1}^* \Delta y_{t-p+1} + \gamma \alpha' y_{t-1} + \varepsilon_t & (2)\\
&= \begin{bmatrix} A_1^* & \ldots & A_{p-1}^* & \gamma \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \alpha' y_{t-1} \end{bmatrix} + \varepsilon_t & (3)\\
&= \Phi Z_{t-1} + \varepsilon_t, & (4)
\end{aligned}$$

where γ and α are full rank matrices of order $n \times q$, q is the rank of the cointegrating space, $-\left(I - \sum_{i=1}^{p} A_i\right) = \gamma\alpha'$, and $A_j^* = -\sum_{i=j+1}^{p} A_i$, $j = 1, \ldots, p-1$. Cointegrating vectors are stacked in α'. Vahid and Engle(1993) show that the VAR may have additional cross-equation restrictions if there are SCCF (or common cycles):

Definition 1 (Vahid and Engle(1993)) *The variables in y_t are said to have SCCF if there are $n - r$ linearly independent vectors, stacked in an $(n-r) \times n$ matrix $\widetilde{\alpha}'$, with the property that:*

$$\begin{aligned}
\underset{(n-r)\times n}{\widetilde{\alpha}'} A_i^* &= 0, \ i = 1, 2, \cdots, p-1, \ \textit{and,} \\
\underset{(n-r)\times n}{\widetilde{\alpha}'} \gamma &= 0.
\end{aligned}$$

Matrix $\widetilde{\alpha}'$ stacks the cofeature vectors, which can be rotated as:

$$\begin{bmatrix} I_{n-r} \\ \widetilde{\alpha}^*_{r\times(n-r)} \end{bmatrix}.$$

Considering rotations of $\widetilde{\alpha}' \Delta y_t$ as $n - r$ equations in a simultaneous-equation system, and completing the system by adding the unconstrained VECM equations for the remaining r series, we obtain,

$$\begin{aligned}
\begin{bmatrix} I_{n-r} & \widetilde{\alpha}^{*\prime} \\ 0 & I_r \end{bmatrix} \Delta y_t &= \begin{bmatrix} 0 & \cdots & 0 & 0 \\ A_1^{**} & \cdots & A_{p-1}^{**} & \gamma^* \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \alpha' y_{t-1} \end{bmatrix} + v_t, & (5)\\
&= \Phi^* Z_{t-1} + v_t, & (6)
\end{aligned}$$

where A_i^{**} and γ^* represent the partitions of A_i^* and γ respectively, corresponding to the bottom r reduced form VECM equations, and $v_t = \begin{bmatrix} I_{n-r} & \widetilde{\alpha}^{*\prime} \\ 0 & I_r \end{bmatrix} \varepsilon_t$. Notice that the matrix Φ^* will be of reduced-rank r, and:

$$\underset{(n-r)\times n}{\widetilde{\alpha}'} \Phi^* = 0.$$

It is easy to show that (5) parsimoniously encompasses (2), since $\begin{bmatrix} I_{n-r} & \widetilde{\alpha}^{*\prime} \\ 0 & I_r \end{bmatrix}$ is invertible, allowing to recover (2) from (5), and the latter has fewer parameters than the former.

Hecq, Palm and Urbain(2005) consider what they call weak-form serial-correlation common features. In this case, only restrictions coming from the short-run dynamics hold, $\widetilde{\alpha}' A_i^* = 0$, $i = 1, 2, \cdots, p-1$, but not $\widetilde{\alpha}'\gamma = 0$.

Definition 2 (Hecq, Palm and Urbain(2005)) *The variables in y_t are said to have SCCF in weak-form if there are $n-r$ linearly independent vectors, stacked in an $(n-r)\times n$ matrix $\widetilde{\alpha}'$, with the property that:*

$$\underset{(n-r)\times n}{\widetilde{\alpha}'} A_i^* = 0, \;\; i = 1, 2, \cdots, p-1.$$

The conditions listed by Vahid and Engle are labelled SCCF in strong-form.

It is straightforward to write the dynamic structure of the system in the case of weak-form SCCF:

$$\begin{bmatrix} I_{n-r} & \widetilde{\alpha}^{*\prime} \\ 0 & I_r \end{bmatrix} \Delta y_t = \begin{bmatrix} 0 & \cdots & 0 & \gamma_1^* \\ A_1^{**} & \cdots & A_{p-1}^{**} & \gamma^* \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \alpha' y_{t-1} \end{bmatrix} + v_t, \tag{7}$$

$$= \Phi^{**} Z_{t-1} + v_t, \tag{8}$$

where $\gamma = \begin{pmatrix} \gamma_1^* \\ \gamma^* \end{pmatrix}$. Notice that Φ^{**} will not be of reduced-rank.

Cointegrated VARs such as (2) can be estimated using OLS with a two-step procedure replacing $\alpha' y_{t-1}$ by $\widehat{\alpha}' y_{t-1}$, where $\widehat{\alpha}'$ stacks super-consistent estimates of cointegrating vectors; see Johansen and Lin and Tsay(1996). Estimation of (5) and (7) is performed by full-information maximum likelihood (FIML), because errors and regressors are correlated. Likelihood-ratio tests can be used to do inference on the number of cofeature vectors $\widetilde{\alpha}'$. These can be based on squared canonical correlations between Δy_t and Z_{t-1}. In previous work, SCCF tests were performed conditional on cointegration using $\widehat{\alpha}' y_{t-1}$. Under correct specification, log likelihood-ratio tests would have a limiting χ^2 distribution; see Vahid and Engle.

2.2 Forecasting

In our discussion about forecasting, we consider the basic VECM (2), with or without the restrictions imposed by SCCF in strong-form. Denote by h the forecasting horizon ($h > 0$) and by $\Delta y_{t+h|t}$ and $y_{t+h|t}$ respectively the linear projections of Δy_{t+h} and y_{t+h} on information dated t and earlier. The h-step ahead forecasts of Δy_t using (2) is:

$$\Delta y_{t+h|t} = A_1^* \Delta y_{t-1+h|t} + \ldots + A_{p-1}^* \Delta y_{t-p+h|t} + \gamma\alpha' y_{t-1+h|t}. \tag{9}$$

As long as long-term constraints (cointegration) hold in the data, there will be linear constraints in forecasting as $h \to \infty$. If we pre-multiply (9) by $(\gamma'\gamma)^{-1}\gamma'$, and take limits as $h \to \infty$, $\Delta y_{t+h|t} \to \mathbb{E}(\Delta y_t) = 0$, and we obtain the well known result that cointegrating linear combinations of level forecasts are colinear,

$$\lim_{h\to\infty} \alpha' y_{t-1+h|t} = 0, \tag{10}$$

although level forecasts of individual series in $y_{t+h|t}$ diverge. Obviously, for any finite h forecasts $\alpha' y_{t-1+h|t}$ will not be colinear. Forecasts $\Delta y_{t+h|t}$ are well defined for all h, but they are not colinear for any h finite.

If short- and long-term constrains (cointegration and strong-form SCCF) hold in the data, not only (10) holds as $h \to \infty$, but forecast $\Delta y_{t+h|t}$ will also be colinear at any finite horizon h. This can be seen by pre-multiplying (9) by $\widetilde{\alpha}'$,

$$\widetilde{\alpha}' \Delta y_{t+h|t} = 0, \tag{11}$$

because $\widetilde{\alpha}' A_i^* = 0$, $i = 1, 2, \cdots, p-1$, and $\widetilde{\alpha}'\gamma = 0$. This point was stressed by Vahid and Issler(2002) to show the importance of SCCF for forecasting with VAR models. We usually build econometric models to forecast at small or medium h, since forecasting uncertainty associated with $\Delta y_{t+h|t}$ gets close to $\mathbb{E}(\Delta y_t \Delta y_t')$ as $h \to \infty$. In these cases, only short-term constraints help, as is the case with SCCF in VAR models.

Suppose that we are not interested in forecasting differences, but levels. One may be tempted to argue that only restrictions on levels matter, such as cointegration. However, this is not true. Suppose we start at $y_t = \mathbb{E}(\Delta y_t) = 0$, and want to forecast y_t h-periods into the future:

$$y_{t+h} = \sum_{i=1}^{h} \Delta y_{t+i}. \tag{12}$$

Its forecast, conditional on information t and earlier is given by:

$$y_{t+h|t} = \sum_{i=1}^{h} \Delta y_{t+i|t}. \tag{13}$$

Pre-multiplying (13) by $\widetilde{\alpha}'$ we obtain,

$$\widetilde{\alpha}' y_{t+h|t} = 0, \tag{14}$$

showing that there is colinearity for $y_{t+h|t}$ at every horizon h.

If instead we impose long-term restrictions coming from cointegration, we obtain:

$$\alpha' y_{t-1+h|t} = \alpha' \sum_{i=1}^{h} \Delta y_{t+h} \neq 0,$$

but,

$$\lim_{h\to\infty} \alpha' y_{t-1+h|t} = 0.$$

3 Model selection criteria and estimation of reduced-rank VAR models

One of the objectives of this paper is to compare the forecasting performance of VAR models when short- and long-term restrictions to the data are considered. To make our results useful to the applied researcher using VAR models for forecasting, we follow the modal strategy in applied work for model selection and estimation. As is well known, most applied research do not consider the presence of SCCF, despite being commonplace testing for cointegration and imposing cointegration restrictions in VAR models used for forecasting. However, when data contain short- and long-term restrictions, ignoring the former may have an impact on the final selection of VAR models, which will be potentially misspecified because the chosen lag length is too short, leading to inconsistent VAR estimates. Of course, this will impact the forecasting performance of VAR models in this context, which is one of the issues we study here.

When dealing with potentially cointegrated VARs, a usual procedure in applied work is to estimate first the long-term coefficient matrix $-\left(I-\sum_{i=1}^{p}A_i\right)=\gamma\alpha'$. The estimate of α' is then used to obtain estimates of short-term coefficient matrices A_i^* and of γ. There is a hierarchy in estimation going from long-term coefficient matrices to short-term coefficient matrices, which typically entails the following steps:

1. Using standard information criteria (AIC, HQ or SC), the lag length of the VAR in levels is chosen for subsequent cointegration analysis.

2. Using the lag length chosen in step 1 above, cointegrating rank and vectors are estimated using the full-information maximum likelihood (FIML) method proposed in Johansen(1989, 1991).

3. Conditional on the results of cointegration analysis, a final VECM is estimated and multi-step ahead forecasts are computed.

According to Johansen(1991), the critical issue of model selection for FIML estimates occurs when selecting the lag length of the VAR. As shown by Gonzalo(1994), the cost of overparameterizing is small. This is what we should expect *a priori*, since estimating a VAR with higher order than necessary only hurts efficiency but not the consistency of parameter estimates. Therefore, the real issue is the cost of underparameterizing, as noted above. In this case, we obtain inconsistent estimates of the VAR coefficient matrices, which yields inconsistent estimates of cointegrating vectors and rank.

Clements and Hendry(1995) and Lin and Tsay(1996) showed that the appropriate choice of the number of unit-roots in the system is critical for out-of-sample forecasting. Moreover, for small and medium sample sizes, Vahid and Issler(2002) showed that, for VAR models with SCCF restrictions and $I(0)$ variables, the performance of information criteria choosing lag length is downward biased toward choosing underparameterized models with too short a lag length. If this result carries through to the case of cointegrated VARs, then steps 1, 2, and 3 above may be a bad strategy for model selection. In this case, given results in Clements and Hendry and Lin and Tsay, it will also be a bad strategy for forecasting, and that is why we worry about it here.

In our context, because we want to design a model selection strategy that avoids the problem of underparameterizing our estimated VAR, we will consider a new strategy for lag-length selection. Instead of using standard information criteria for choosing the lag length of the VAR in levels, we follow Vahid and Issler choosing simultaneously lag order p and the number of common cycles r (i.e. the rank of Φ^*), by minimizing the following criteria[1],

$$AIC(p,r) = \sum_{i=n-r+1}^{n} \ln\left(1-\lambda_i\left(p\right)\right) + \frac{2}{T} \times r \times \left(n\left(p-1\right)+q+n-r\right) \quad (15)$$

$$HQ\left(p,r\right) = \sum_{i=n-r+1}^{n} \ln\left(1-\lambda_i\left(p\right)\right) + \frac{2\ln\ln T}{T} \times r \times \left(n\left(p-1\right)+q+n-r\right) \quad (16)$$

$$SC\left(p,r\right) = \sum_{i=n-r+1}^{n} \ln\left(1-\lambda_i\left(p\right)\right) + \frac{\ln T}{T} \times r \times \left(n\left(p-1\right)+q+n-r\right), \quad (17)$$

where n is the dimension of the (number of series in the) system, r is the rank of the VEC model, $(p-1)$ is the number of lagged differences in the VECM, T is the number of observations, and λ_i are the sample squared canonical correlations between Δy_t and the set of regressors Z_{t-1}.

It is obvious from (15)-(17) that all these information criteria depend on q as well. Therefore, we need to design a strategy for setting q in selecting (p,r). Here we use the idea in Hecq, Palm and Urbain(2005) of SCCF in weak form, where $\widetilde{\alpha}'A_i^* = 0$, $i = 1, 2, \cdots, p-1$, but $\widetilde{\alpha}'\gamma = 0$ does not hold. Notice that cointegration implies rank reduction for γ, because γ is a full-rank matrix of order $n \times q$, and $q < n$. Hence, not imposing any cointegration constraints on the rank of γ is equivalent to consider $q = n$. In this case, γ is square and invertible. Hence, $\widetilde{\alpha}'\gamma = 0$ only accepts a trivial solution for $\widetilde{\alpha}'$, and $\widetilde{\alpha}'\gamma = 0$ does not hold. In selecting (p,r), when $q = n$, we do not impose $\widetilde{\alpha}'\gamma = 0$ by construction, although $\widetilde{\alpha}'A_i^* = 0$, $i = 1, 2, \cdots, p-1$, may hold, depending on the final choice of (p,r). Therefore we implicitly use the idea of weak-from SCCF put forth by Hecq, Palm and Urbain.

Once p and r are chosen, we propose estimating the short-run dynamic matrices A_i^* without imposing any constraints on cointegrating rank q. We do this by running a regression of Δy_t and of Δy_{t-i}, $i = 1, 2, \cdots, p-1$, on y_{t-1}, respectively, saving the respective set of residuals. We then run a reduced-rank regression of the first set of residuals on the second, estimating the short-run dynamic matrices A_i^*, $i = 1, \ldots, p-1$, on a first pass. These matrices will all have rank r. Using these first-pass short-run coefficient matrices, we then estimate $-\left(I - \sum_{i=1}^{p} A_i\right)$, without imposing any rank restrictions on it. This allows getting second-pass estimates of A_i^*, $i = 1, \ldots, p-1$, since we can now compute residuals of Δy_t and of Δy_{t-i}, $i = 1, 2, \cdots, p-1$, on $-\widetilde{\left(I - \sum_{i=1}^{p} A_i\right)} y_{t-1}$, respectively, where the latter is the first-pass estimate of the long-run matrix $-\left(I - \sum_{i=1}^{p} A_i\right)$. These second-pass matrices will all have rank r as well. This algorithm iterates until convergence.

[1] When variables are not cointegrated $q = 0$, and these criteria are the same as those suggested in Lütkepohl(1993, p. 202).

Then, we can compute the canonical correlations between Δy_t and $\widehat{A_1^*}\,\Delta y_{t-1} + \ldots + \widehat{A_{p-1}^*}\,\Delta y_{t-p+1} - \widehat{\left(I - \sum_{i=1}^{p} A_i\right)} y_{t-1}$, where hats denote estimates obtained after convergence of the algorithm. At this stage, it is possible to test for cointegration, determining the rank of $-\left(I - \sum_{i=1}^{p} A_i\right)$ and the corresponding cointegrating vectors using Johansen's method (Trace test, at 5% significance).

At the end, this algorithm produces a choice of p, r, and q, which can be used for estimation and forecasting using a VECM with short- and long-term restrictions. Its forecasting performance can be compared with that of a regular VECM obtained from the first procedure described in this section (currently the most used in applied research), where SCCF restrictions are ignored in every modelling stage. Notice that there is the risk of misspecification when both procedures are used. For the first procedure, the critical issue for misspecification is the lag-length selection by standard information criteria, as pointed out by Vahid and Issler. For the second, although lag-length selection may still generate a misspecified model, this risk is reduced. However, there is the potential for misspecification in the choice of (p, r) and later in the choice of q.

The new algorithm proposed here inverts the hierarchy in estimating short- and long-term restrictions in VAR models. Because cointegrating vectors are super-consistent, the usual practice in the literature is to first estimate cointegrating vectors and cointegrating rank. Using super-consistent estimates of cointegrating vectors, and an estimate of the cointegrating rank, VECM estimation is performed conditional on them. Under correct specification, short-term coefficient matrices will converge at rate $\sqrt{T}$. Our procedure first estimates short-term coefficient matrices after using preferred lag-length selection criteria. This reduces the chance of choosing too small a lag length for the VAR. Based on these short-term estimates, and on unrestricted long-term estimates of $-\left(I - \sum_{i=1}^{p} A_i\right)$, we finally test for cointegration. Hence, long-term coefficients are a function of short-term coefficient matrices estimates, inverting the usual practice in the literature. This, we hope, will open up the discussion on joint estimation of short- and long-term coefficient matrices estimates.

4 Monte-Carlo design for VARs with short- and long-term restrictions

One of the critical issues in any Monte-Carlo study is that of diversity of Data Generating Processes (DGPs), which allows sampling a wide spectrum of the parameter space. One of the limitations in our context is that the VAR contains short- and long-term restrictions, which must hold simultaneously in every DGP. If we fix the cointegrating and the cofeature vectors, α and $\widetilde{\alpha}$ respectively, the VAR coefficients must then obey simultaneously $-\left(I - \sum_{i=1}^{p} A_i\right) = \gamma\alpha'$, $\widetilde{\alpha}' A_i^* = 0$, $i = 1, 2, \cdots, p-1$, and $\widetilde{\alpha}'\gamma = 0$. Also, the eigenvalues of the companion matrix of the VAR have all to be on or inside the unit circle (or all the eigenvalues of the companion matrix of the VECM have to be inside the unit circle). The simulation exercise would be very time consuming if we simply fix α and $\widetilde{\alpha}$ and then randomly select A_i, $i = 1, 2, \cdots, p$, verifying whether these restrictions hold for every one of these choices. Although this procedure will certainly sample a wide spectrum of the parameter space, with current PC technology, it will take more than 50 years to be completed, and is obviously ruled out. There are two alternatives. The first is to fix α and $\widetilde{\alpha}$ and then solve analytically what are restrictions the elements of A_i, $i = 1, 2, \cdots, p$, must obey in order for the eigenvalues of the companion matrix of the VECM to be inside the unit circle. This will somehow limit the search on the parameter space, but is feasible. The only problem is that the number of series n and the number of lags p cannot be too big, otherwise finding the analytical solution becomes a time-consuming problem. This is the main procedure used in our simulation study. The alternative is to fix the restricted VECM coefficients (equation (5)), varying them slightly, verifying whether the eigenvalues of the companion matrix of the VECM are inside the unit circle in every case. This procedure imposes greater limits on sampled parameter space, but is the most practical one, and it is used when we investigate the performance of the new algorithm proposed above.

To make the Monte-Carlo simulation manageable, we propose using as DGP a three-dimensional VAR, i.e., $n = 3$. Models that consider the real side of the economy are often three-dimensional. For example, King et al. (1991) estimate a VAR including output, consumption, and investment in order to test the real-business-cycle model of King, Plosser and Rebelo (1988). The first parameter we set in the Monte-Carlo design is the lag length $p = 3$ of the VAR (the lag length of the VECM is $p - 1 = 2$). This choice allows either under- or over-parameterization of the VAR model, which is an important ingredient of any VAR Monte-Carlo study as stressed by Vahid and Issler(2002):

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + \varepsilon_t. \tag{18}$$

Next, we set the number of cointegrating vectors to one, i.e., $q = 1$, and the number of cofeature vectors to two, i.e., $n - r = 2$, or $r = 1$. The cointegrating and cofeature vectors are respectively:

$$\alpha = \begin{bmatrix} 1.0 \\ 0.2 \\ -1.0 \end{bmatrix} \text{ and } \widetilde{\alpha} = \begin{bmatrix} 1.0 & 0.1 \\ 0.0 & 1.0 \\ 0.5 & -0.5 \end{bmatrix}. \tag{19}$$

Conditional on these values, we then choose the number of free parameters remaining in the coefficient matrices A_1, A_2, and A_3 in order to keep all the eigenvalues of companion matrix of the VECM inside the unit circle. Appendix A contains a detailed discussion of the final choice of these free parameters, including analytical solutions.

The final number of DGPs in which $q = 1$, and $r = 2$, satisfying (19), with eigenvalues of the companion matrix inside the unit circle, was set equal to 100. For each of these 100 DGPs, we generated 1,000 samples of y_t's, by sampling random series ε_t's. Each of these 1,000 samples had 1,000 observations. However, in all cases, to reduce the impact of initial values on simulated series, we only used the last $T = 100$ or $T = 200$ observations in running regressions. Therefore, our final results will be based on 1,000 samples of 100 different DGPs – a total of 100,000 different samples – of either $T = 100$ or $T = 200$ observations.

As discussed in Vahid and Issler, it is worth sorting results by signal-to-noise ratio (or system R^2 measures). Here, we selected two different set of parameters with the following characteristics: the first has the median of the system R^2 measure between 0.4 and 0.5, with 3% larger than 0.6 and none greater than 0.7. The second has the median of the system R^2 between 0.7 and 0.8, with 22% larger than 0.8, none greater than 0.9, and none smaller than 0.7.

The Monte-Carlo procedure can be summarized as follows. Using each of our 100 DGPs, we generated 1,000 samples (once with 100, and again with 200 observations). Then, we recorded the lag length chosen by traditional (full-rank) information criteria, labelled $IC(p)$: $AIC(p)$, $HQ(p)$ and $SC(p)$[2], and the corresponding lag length chosen by alternative information criteria, labelled $IC(p,r)$: $AIC(p,r)$, $HQ(p,r)$ and $SC(p,r)$, in (15)-(17), with $q=3$.

For choices made using $IC(p)$ we used Johansen's(1989, 1991) trace test at 5% to choose q and then estimated a VECM with no SCCF restrictions. Their out-of-sample forecasting accuracy measures were recorded up to 16 periods ahead. For choices made using $IC(p,r)$, we used the algorithm described in detail in the last section to obtain a triplet (p,q,r) in each case, with a resulting VECM estimated using SCCF restrictions. Their respective out-of-sample forecasting accuracy measures were recorded up to 16 periods ahead. Out-of-sample forecasting accuracy measures were then compared for these two types of VAR estimates.

4.1 Measuring forecast accuracy

The loss functions used here to compute forecasting accuracy are a blend of tradition, such as the determinant of the mean-squared forecast error matrix at different horizons ($|MSFE|$) and the trace of the mean-squared forecast error matrix ($TMSFE$), and of modern loss functions that are invariant to linear transformation of forecasts, such as Clements and Hendry's(1993) *generalized forecast error second moment* (*GFESM)*. The excellent discussion in Lin and Tsay(1996) justifies the use of these two types of loss functions in measuring forecast accuracy.

There is one complication associated with simulating 100 different DGPs. Simple averaging across different DGPs is not appropriate, because the forecast errors of different DGPs do not have identical variance-covariance matrices. Lütkepohl(1985) normalizes the forecast errors by their true variance-covariance matrix in each case to get i.i.d. observations. Unfortunately, this would be a very time consuming procedure for a measure like *GFESM,* which involves stacked errors over many horizons. Instead, for each information criterium, we calculate the percentage gain in forecasting measures, comparing the full-rank models selected by $IC(p)$, with the reduced-rank models chosen by $IC(p,r)$. The percentage gain is computed using natural logs of ratios of respective loss functions, since

[2] Formulas for $IC(p)$ are as follows:

$$AIC(p,r) = \sum_{i=1}^{n} \ln(1-\lambda_i(p)) + \frac{2}{T} \times n^2 p, \tag{20}$$

$$HQ(p,r) = \sum_{i=1}^{n} \ln(1-\lambda_i(p)) + \frac{2\ln\ln T}{T} \times n^2 p, \tag{21}$$

$$SC(p,r) = \sum_{i=n-r+1}^{n} \ln(1-\lambda_i(p)) + \frac{\ln T}{T} \times n^2 p. \tag{22}$$

this implies symmetry of results for gains and losses. This procedure is done at every iteration, for every DGP, and the final results are then averaged.

5 Monte-Carlo simulation results

5.1 Known lag order and cointegrating and cofeature ranks

In order to establish a benchmark, we first examine what are the potential gains of imposing SCCF restrictions when all chosen models are correctly specified, i.e., we know the lag order, cointegrating rank, and rank of Φ^*: (p, q, r). Cointegration vectors are estimated by FIML as proposed by Johansen(1988, 1991) for models that ignore SCCF restrictions. Then, conditional on $\widehat{\alpha}'$, a VECM is estimated by OLS, equation-by-equation. For models that take into account SCCF restrictions, cointegrating vectors are estimated by the iterative procedure discussed in Section 3. We also include a unrestricted VAR in levels in the analysis, which is estimated by OLS, equation-by-equation.

Simulation results are presented in Table 1, allowing the following conclusions. First, the result in Engle and Yoo(1987) that the VAR in levels outforecasts the VECM at short horizons only happens when the system R^2 is high and does not seem to hold as a general result for vector error-correction models. Indeed, it seems that the rule is that the VECM outforecasts the unrestricted VAR in short and long horizons. Hence, the strategy of presenting results according to the value of the system R^2 has an important payoff – changing the conventional wisdom with regard to forecasting with VECMs at short horizons. Second, the restricted VECM outforecasts the unrestricted VECM almost everywhere. Percentage gains can be higher than 100% in some short-horizon cases, although gains tend to decline as the horizon increases. This final result is what we should expect *a priori*: a constrained model should forecast better than an unrestricted model whenever these constraints are true, which is the case here. The only point of interest is: by how much? The benchmark numbers obtained here in favor of the constrained model are encouraging. Comparisons of the restricted VECM with the unrestricted VAR come out in favor of the former almost everywhere as well, the exception being when the system R^2 is high and the horizon is shortest. We next consider the possibility of model misspecification in our simulation study.

5.2 Selection of lag and rank order and number of cointegrating vectors

We now examine the performance of standard information criteria, followed by Johansen's(1988, 1991) cointegrating test in selecting the number of lags of the VAR in levels p and the number of cointegrating vectors q when the DGP is a $VAR(3)$ with $r = 1$, and $q = 1$. Table 2 shows the frequency of choice of p and q in 1,000 simulations of 100 trivariate VARs with a low system R^2. Table 3 shows the same when systems with high R^2s are considered.

We can draw the following conclusions from Tables 2 and 3. First, the total frequency in which the true lag length $p = 3$ is selected (adding up row 3 for all three $IC(p)$) varies very little with respect to the system R^2, being about 50%, 20%, and 2%, for $AIC(p)$, $HQ(p)$ and $SC(p)$, respectively, when $T = 100$, and about 80%, 50% and 10%, when $T = 200$. Although the true lag length and number of cointegrating vectors of the DGP is $(3, 1)$, it is never the mode in Table 2, and is rarely the mode in Table 3. The performance of using $IC(p)$ to select p and q is discouraging for $HQ(p)$ and $SC(p)$: even when $T = 200$, and system R^2s are high, $(p, q) = (3, 1)$ is selected only 44.22% and 8.31% of times, respectively. Despite that, the frequency in which $(3, 1)$ is selected rises in both directions (sample size and R^2 measure) as expected. Second, the $AIC(p)$ criterium selects more frequently VARs where the number of lags is overestimated than the other criteria. This result is very similar to the one reported in Vahid and Issler(2002). Also, in Tables 2 and 3, $AIC(p)$ chooses the correct (p, q) pairs more often than the other two criteria. The modal choice of the $SC(p)$ criterium is a $VAR(2)$ with one cointegrating vector even when the number of observations is 200 and the system R^2 is high.

Table 4 shows the frequency of lag-rank-cointegrating vectors (p, r, q) selection in 1,000 simulations of 100 trivariate $VAR(3)$ with $r = 1$, one cointegrating vector and low system R^2. There are two steps in this selection: (i) chose the number of lags and rank simultaneously by $AIC(p, r)$, $HQ(p, r)$ and $SC(p, r)$; and (ii) use this information to perform Johansen's cointegrating test to select the number of cointegrating vectors after applying the algorithm described in Section 3. Table 5 shows the analogous results when a high system R^2 is considered.

The following conclusions emerge from analyzing Tables 4 and 5. First, the total frequency in which the true lag length (2 lags in differences) is selected (adding up row 2 for all three $IC(p, r)$) varies very little with respect to the R^2 of the system: it is about 60%, 67%, and 48%, for $AIC(p, r)$, $HQ(p, r)$ and $SC(p, r)$, respectively, when $T = 100$, and about 78%, 91% and 78%, when $T = 200$. These results are equal or better than those of standard information criteria $IC(p)$. The mode for the selected lag length is always the correct one – 2 lags in differences – except for $SC(p, r)$. Second, the $AIC(p, r)$ criterium selects more frequently VARs where the number of lags is overestimated than the other criteria. In both tables, $HQ(p, r)$ chooses the correct (p, q, r) triplets more often than the other two criteria, with $SC(p, r)$ in second place. For high system R^2s, the modal choice of $HQ(p, r)$ criterium is the correct one with a relatively high frequency. Third, for low system R^2s, and sample size $T = 100$, the true model is selected about 8%, 10%, and 8% of times using respectively $AIC(p, r)$, $HQ(p, r)$ and $SC(p, r)$. For high system R^2 measures, and sample size $T = 200$, these same frequencies are about 78%, 91% and 78%.

When we compare the results in Tables 2 and 3 with those of Tables 4 and 5, it becomes apparent that using information criteria in selecting (p, r) jointly fares much better than selecting p alone. In Tables 1 and 2 the true model with $p = 3$ and $q = 1$ is selected by $AIC(p)$, $HQ(p)$ and $SC(p)$ respectively with frequency ranging from 8%-71%, 4%-44%, and 1%-8%. The equivalent figures in Tables 4 and 5 for $AIC(p, r)$, $HQ(p, r)$ and $SC(p, r)$ are: 8%-66%, 10%-78%, and 8%-68%. The only instance when $IC(p)$ performs better than $IC(p, r)$ is when $AIC(\cdot)$ is used. Even then, differences are rather small. The direction in which $IC(p)$ performs badly is also worrisome, since the rule is for $IC(p)$ to choose too small a lag length, leading to models with inconsistent estimates.

Overall, the results in Tables 2 to 5 confirm that, when data have cointegration and SCCF restrictions, ignoring the latter has a high cost in terms of model selection. This happens because applied researchers usually select lag length of VAR models using standard information criteria. However, as our simulation study confirms, SCCF restrictions leads to selected VAR models with a number of lags that are generally smaller than true ones, generating inconsistent estimates of the number of lags and of cointegrating vectors. As discussed in previous theoretical and applied studies, if data have unit roots, selecting them appropriately is critical for consistent estimation and forecasting efficiency; see Johansen(1991), Gonzalo(1994), Clements and Hendry(1995) and Lin and Tsay(1996). However, the latter is jeopardized by the traditional way of model selection under cointegration and SCCF restrictions, which calls for a different approach – selecting lag and rank simultaneously using $IC(p,r)$.

5.3 Forecasts

In this section, we compare the forecasting performance of three different estimation procedures for VARs containing short- and log-term restrictions. The first is to estimate the system in levels with no restrictions using OLS, equation-by equation. Lag length is selected using standard information criteria. The second is to select lag length using standard information criteria, later imposing long-term restrictions arising from cointegration, and estimating a VECM. The third is to impose simultaneously short- and log-term restrictions in VECM estimation using the algorithm described in Section 3, after selecting lag length and the rank of Φ^* simultaneously.

Tables 6 to 11 exhibit pairwise percentage improvement in forecasting for different models, when we consider systems with low and high R^2s. Numbers in boldface denote the best information criterium for model selection when we use as loss function the trace of the mean-squared forecast error (lowest value), while underlined numbers denote the worst information criterium for model selection.

When we compare reduced-rank VECMs with unrestricted VECMs, we observe large percentage improvements at short horizons. These are really impressive – as big as 138% – when system R^2s are high and $T = 100$. Even when R^2s are low and $T = 100$, we observe percentage improvements ranging from 7%-20% for reduced-rank VECMs over VECMs.

When we compare the VECM with the unrestricted VAR we observe that the former fares better in general, although when system R^2s are high, and the horizon is low, the unrestricted VAR performs better. Unless one uses GFESM, as the horizon increases, the percentage gains of the VECM over the unrestricted VAR increases, reaching more than 800% in some cases. Of course, this is a consequence of imposing long-term restrictions as stressed by Engle and Yoo(1987). In Tables 8 and 9 there is a large forecasting improvement over all horizons when the R^2 measure is low, opposite to the findings in Engle and Yoo. Of course, this is reverted when R^2s are high, showing that their results hold only as a special case as discussed in Section 5.1.

It is hard to use the results in Tables 6 to 11 to choose overall the best information criterium to use for lag-length selection. However, if one suspects that the data has SCCF and cointegration restrictions, then the HQ criterium should be used if the system R^2 is low, while AIC should be used if the system R^2 is high. The worst criterium to use in the first case is AIC, while in the second case it is SC. Notice that using HQ avoids selecting the worst models for forecasting, which may be a deciding factor in its favor. If one is considering only long-term restrictions AIC should be avoided, since it produces the largest loss overall. A strong candidate to be used here is the HQ criterium, especially if the system R^2 is low.

Tables 10 and 11 show the total improvement of imposing short- and log-term restrictions – comparing the reduced-rank VECM with the unrestricted VAR. Gains are non-trivial at all horizons, reaching over 800% in some cases. The exception is when the system R^2 measures are high and the horizons are lowest, although as it increases gains can reach more than 600%.

Overall, it seems that considering appropriately constrained models helps in forecasting. These gains can be important in short horizons, especially when high system R^2s are considered. As seen above, the gains in imposing long-term restrictions only materialize as the horizon increases, while those of imposing short-term restrictions are usually obtained at short-horizons. Because forecasting uncertainty at long horizons can be large, time-series models are generally most useful for forecasting at short horizons. Hence, imposing short-term constraints are a way of improving the effectiveness of time-series models at horizons where they are most useful. This is one of the main points of our paper, which is not stressed very often in the forecasting literature. The other is the comparison of the gains arising from imposing short- and long-term restrictions in VAR estimation, which came always in favor of imposing short-term restrictions, with or without considering the possibility of misspecified models.

6 Estimation performance of the new estimation algorithm

Our last investigation is on the performance of the new estimation algorithm, described in detail in Section 3. To save space, our simulation study will not be as broad as the one presented above in the forecasting experiment. Instead of fixing α and $\widetilde{\alpha}$, solving analytically for A_i, $i = 1, 2, \cdots, p$, imposing the restriction that all the eigenvalues of the companion matrix of the VECM to be inside the unit circle, we fixed the restricted VECM coefficients (equation (5)), varying them slightly, verifying whether the eigenvalues of the companion matrix of the VECM are inside the unit circle in every case. To limit further the scope of our simulation study, we will assume that the lag length of the VAR (p), the cointegrating rank (q) and the coefeature rank ($n - r$) are known with certainty. Therefore, results here do not allow for model misspecification, and are consistent with those presented in Table 1.

We will examine the mean-squared error (MSE) in estimating short- and long-term coefficients in the VAR. In all DGPs, the VAR in levels is assumed to be of order 2, i.e., a VECM of order 1. The number of variables in the VAR was set equal to 3, i.e., $n = 3$, and we have varied q and r with all possible combinations: $q = 1$ and $r = 1$, $q = 1$ and $r = 2$, and $q = 2$ and $r = 2$. In order to summarize MSE information we stack respectively short-term coefficients, long-term coefficients, and all coefficients in a vector, computing the determinant of the MSE matrix $|MSE|$ as well as its trace ($TMSE$). We also report MSEs of individual coefficients.

First, MSE results were computed for VECMs, with cointegrating vectors estimated by FIML using Johansen's(1988, 1991) technique, imposing the true value of q. Other coefficients were estimated by OLS, equation-by-equation. No reduced-rank structure for Φ^* is imposed in this case. Second, MSE results were computed imposing a reduced-rank structure for Φ^*, using the algorithm described in detail in Section 3: first estimate short-term coefficients imposing $rank\,(\Phi^*) = r$, later estimating $-\left(I - \sum_{i=1}^{p} A_i\right)$ using short-term estimates, iterating until convergence. After convergence, we impose the true value of q in estimating the cointegrating vector and the components of γ. We report the percentage gains in MSE measures of the second procedure over the first for every DGP, finally averaging across all DGPs. We have 100 different DGPs, each with 1,000 different simulations. Results are grouped by system R^2s.

Table 12 (a and b) presents results for $q = 1$ and $r = 1$. Here, rank reduction is the highest possible, leading to a relatively high payoff of imposing rank restrictions in Φ^*. As expected, estimation of short-term coefficients largely benefit from correctly imposing $r = 1$. For high system R^2s, percentage gains in $TMSE$ are higher than 50% when $T = 100$, and higher than 60% when $T = 200$. For low system R^2s, these numbers are about 40%. What is interesting about Table 12, is the percentage gain on the $TMSE$ of long-term coefficients: with high system R^2s, for $T = 100$ or $T = 200$ they are about 15%, showing that long-term coefficient estimation can also benefit from imposing valid short-term restrictions. It also interesting to observe that γ estimates benefit the most, while for estimates of α' the gains are more modest. This is a consequence of the fact that estimates of α' using FIML are super-consistent. When system R^2s are low, there is a loss of about 7% in estimating α_1 for $T = 100$, which reverts to a 4.86% gain for $T = 200$. Notice that results for $|MSE|$ compound variance and covariance gains, leading to much higher percentage gains.

Table 13 (a and b) presents results for $q = 1$ and $r = 2$, while Table 14 (a and b) presents results for $q = 2$ and $r = 2$. We should expect a smaller percentage gain because there are not as many restrictions in Φ^* as in the case where $r = 1$. Still, in Table 13, long-term coefficient estimates show an improvement in $TMSE$ of 8.5% for $T = 100$, when system R^2s are high, falling to 6.5%, when system R^2s are low. Improvement in short-term coefficient estimates are also sizable, despite the fact that we observe a few cases where there is a loss in using the algorithm. In Table 14, we present long-term coefficients in canonical echelon form, since the cointegrating rank is now 2. In this case, improvement in $TMSE$ for long-term coefficients are 19.79% for $T = 100$, when system R^2s are high, and 20.55%, when system R^2s are low. Gains for short run coefficients measured by $TMSE$ can reach more than 200% when system R^2s are low, and more than 300% when system R^2s are high.

Overall, it seems that the new algorithm performs very well not only in estimating short-term coefficients but also in estimating long-term coefficients. This is a consequence of the fact that long-term coefficients are a cumulation of short-term coefficients. Therefore, gains in estimating the latter can translate into gains in estimating the former. Further research should focus on the long due issue of joint estimation of short- and long-term parameter estimation when VAR models are subject to short- and long-term restrictions.

7 Conclusion

In empirical studies, common-cyclical features have been shown to exist for a variety of macroeconomic data sets: Campbell and Mankiw (1989) find a common cycle between consumption and income for most G-7 countries, Engle and Kozicki (1993) find common international cycles in GNP data for OECD countries, Vahid and Engle(1993) and Issler and Vahid (2001) find common cycles for macroeconomic aggregates, Engle and Issler (1995) and Carlino and Sill (1998) find common cycles for sectoral and regional outputs respectively, Issler and Vahid(2005) find common cycles in U.S. coincident series, and

Hecq, Palm and Urbain(2005) and Hecq(2005) find common cycles for Latin American GDPs. In all these articles, the dynamic representation of the data was a cointegrated VAR. A natural question which then arises is the following: what are the consequences of imposing short- and long-term restrictions in estimation and forecasting if a VAR is used as the dynamic representation of the data? The objective of our paper is to answer this question, hoping that the answer will be useful for model building.

In order to investigate these issues, we take VAR models, which have become the "working horses" for macroeconometric studies, and investigate the potential gains for forecasting and estimation uncertainty of imposing short- and long-term restrictions arising from the existence of common cycles and cointegration. The environment of our study is that of simulation, and our exercise is devised in such a way that the results are applicable to an applied researcher which has access to a relatively small number of time-series of observations, making parsimony a critical issue in model building.

In model selection and forecasting, we compare the behavior of the two strategies. The first is widely employed in applied work: VAR order is selected by standard information criteria $IC\,(p)$ and later used in testing for cointegration. The existence of common cycles is completely ignored and forecasting is performed with a standard VECM. The second takes into account short- and long-term restrictions in a novel way. In a first step, information criteria $IC\,(p,r)$ are used to choose lag length and rank order simultaneously. Next, weak-form SCCF restrictions on VECM coefficient matrices are used in devising a new algorithm for joint estimation of short- and long-term parameters of the VAR. Forecasting is based on a final model taking into account these two sets of restrictions. We also compare the estimation performance of these two strategies in model building, where short- and long-term parameters are considered separately.

First, our results confirm that, when data have cointegration and SCCF restrictions, ignoring the latter has a high cost in terms of model selection. This happens because $IC\,(p)$ chooses too frequently inconsistent models, with too small a lag length. Choosing lag and rank simultaneously using $IC\,(p,r)$ has a superior performance in this case, reducing drastically the frequency in which inconsistent models are selected. Second, the superior performance of $IC\,(p,r)$ over $IC\,(p)$ translates into a superior forecasting performance of the restricted VECM over the VECM, with considerable gains in some cases. Our conclusion is that, overall, there is a relatively large forecasting improvement for small horizons when SCCF restrictions are accounted for. Results for systems with high R^2 measures are really impressive. Third, the new algorithm proposed here fares very well in terms of parameter estimation, even when we consider the estimation of long-term parameters.

These and previous results on SCCF restrictions in VAR analysis call for a change in focus in the literature on forecasting with these models. While in the past there has been a considerable effort examining the importance of cointegration restrictions in VAR models, there has been very little work examining the importance of common-cyclical features in VAR analysis. We hope that our results and others will help changing the focus of the literature towards SCCF restrictions.

References

[1] Ahn, S.K. and G.C. Reinsel (1988), "Nested reduced-rank autoregressive models for multiple time series", *Journal of the American Statistical Association*, 83, 849-856.

[2] Beveridge, S. and C.R. Nelson (1981), "A New Approach to Decomposition of Economic Time Series into a Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle", *Journal of Monetary Economics*, vol. 7, pp. 151-174.

[3] Campbell, J.Y. and Mankiw, N.G.(1989), "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," *NBER Macroeconomics Annual.*

[4] Carlino, G. and K. Sill (1998), "Common trends and common cycles in regional percapita incomes", Working Paper, Federal Reserve Bank of Philadelphia.

[5] Cheung, Y.W. and Lai, K.S. (1993), "Finite-samples sizes of johansen likelihood ratio tests for cointegration," *Oxford Bulletin of Economics and Statistics*, 55 (3): 313-328.

[6] Christoffersen, P.F. and Diebold, F.X. (1998), "Cointegration and long-horizon forecasting," *Journal of Business and Economic Statistics*, vol. 16, 4, pp. 450-458.

[7] Clements, M.P. and D.F. Hendry (1993), "On the limitations of comparing mean squared forecast errors," *Journal of Forecasting*, 12, 617-637 (with discussions).

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

[9] Diebold, F.X. and Kilian L. (2001), "Measuring predictability: Theory and macroeconomic applications," *Journal of Applied Econometrics*, 16 (6): 657-669.

[10] Engle, R.F. and C.W.J. Granger (1987), "Cointegration and error correction: Representation, estimation and testing", *Econometrica*, 55, 251-276.

[11] Engle, R.F. and J.V. Issler (1995), "Estimating common sectoral cycles", *Journal of Monetary Economics*, 35, 83-113.

[12] Engle, R.F. and Kozicki, S.(1993), "Testing for Common Features," Journal of Business and Economic Statistics, vol. 11, pp. 369-395, with discussions.

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

[14] Gonzalo, J (1994), "Five alternative methods of estimating long-run equilibrium relationships," *Journal of Econometrics*, 60 (1-2): 203-233.

[15] Granger, C.W.J.(1981), "Some properties of time series data and their use in econometric model specification", *Journal of Econometrics*, 16, pp. 121-130.

[16] Hecq, A. (2005), "Common trends and common cycles in Latin America: a 2-step vs. an iterative approach." Working paper: University of Maastricht.

[17] Hecq, A., Palm, F.C., Urbain, J.-P. (2005), "Testing for common cyclical features in VAR models with cointegration," forthcoming in the *Annals Issue of the Journal of Econometrics* on "Common Features."

[18] Hoffman, D.L. and Rasche, R.H. (1996), "Assessing forecast performance in a cointegrated system," *Journal of Applied Econometrics*, 11 (5): 495-517.

[19] Issler, J.V. and Vahid, F. (2001), "Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates," *Journal of Monetary Economics,* 47, 449-475.

[20] Issler, J.V. and Vahid, F. (2005), "The Missing Link: Using the NBER Recession Indicator to Construct Coincident and Leading Indices of Economic Activity," forthcoming in the *Annals Issue of the Journal of Econometrics* on "Common Features."

[21] Johansen, S.(1988), "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, vol. 12, pp. 231-254.

[22] Johansen, S.(1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, vol. 59, pp. 1551-1580.

[23] King, R., C.I. Plosser, J.H. Stock and M.W. Watson (1991), "Stochastic Trends and Economic Fluctuations", *American Economic Review*, 81, 819-840.

[24] King, R.G., C.I. Plosser and S. Rebelo (1988), "Production, Growth and Business Cycles. II. New Directions," *Journal of Monetary Economics*, vol. 21, pp. 309-341.

[25] Lin, J.L. and R.S. Tsay (1996), "Cointegration constraints and forecasting: An empirical examination", *Journal of Applied Econometrics,* 11, 519-538.

[26] Lütkepohl, H. (1985), "Comparison of criteria for estimating the order of a vector autoregressive process", *Journal of Time Series Analysis*, 6, 35-52.

[27] Stock, J.H. and M.W. Watson (1988), "Testing for common trends", *Journal of the American Statistical Association*, 83, 1097-1107.

[28] Silverstovs, B., Engsted T. and Haldrup N. (2004), "Long-run forecasting in multicointegrated systems," *Journal of Forecasting*, 23 (5): 315-335.

[29] Stock, J. and Watson, M.(1989) "New Indexes of Leading and Coincident Economic Indicators", *NBER Macroeconomics Annual*, 351-95.

[30] Tiao, G.C. and R.S. Tsay (1989), "Model specification in multivariate time series (with discussions)", *Journal of the Royal Statistical Society, Series B,* 51, 157-213.

[31] Tso, M.K-S. (1981), "Reduced rank regression and Canonical analysis" *Journal of the Royal Statistical Society, Series B,* 43, 183-189.

[32] Vahid, F. (1999), "A property of the companion matrix of a reduced rank VAR", *Econometric Theory*, 15, 787-788.

[33] Vahid, F. and R.F. Engle (1993), "Common trends and common cycles", *Journal of Applied Econometrics*, 8, 341-360.

[34] Vahid, F. and Engle, R. (1997), "Codependent cycles," *Journal of Econometrics*, 80 (2): 199-221.

[35] Vahid, F. and Issler, J.V. (2002), "The importance of common cyclical features in VAR analysis: A Monte Carlo study," *Journal of Econometrics*, 109 (2): 341-363.

[36] Wang, Z.J. and Besseler D.A. (2004), "Forecasting of multivariate time series models with full and reduced rank: an empirical examination," *International Journal of Forecasting*, 20 (4): 683-695.

A VAR restrictions for the DGPs

Consider the following $VAR(3)$ in levels:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + \varepsilon_t,$$

which can be rewritten as the following $VAR(1)$ process,

$$\begin{bmatrix} \Delta y_t \\ \Delta y_{t-1} \\ \alpha' y_t \end{bmatrix} = \begin{bmatrix} -(A_2+A_3) & -A_3 & \gamma \\ I_3 & 0 & 0 \\ -\alpha'(A_2+A_3) & -\alpha' A_3 & \alpha'\gamma+1 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta y_{t-2} \\ \alpha' y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \alpha'\varepsilon_t \end{bmatrix}, \tag{23}$$

where, $\gamma\alpha' = (A_1 + A_2 + A_3 - I_3)$, $\widetilde{\alpha} = \begin{bmatrix} \widetilde{\alpha}_{11} & \widetilde{\alpha}_{12} \\ \widetilde{\alpha}_{21} & \widetilde{\alpha}_{22} \\ \widetilde{\alpha}_{31} & \widetilde{\alpha}_{32} \end{bmatrix}$, $\alpha = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \end{bmatrix}$, $\gamma = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{bmatrix}$, $A_2 = \begin{bmatrix} a^2_{11} & a^2_{12} & a^2_{13} \\ a^2_{21} & a^2_{22} & a^2_{23} \\ a^2_{31} & a^2_{32} & a^2_{33} \end{bmatrix}$ and $A_3 = \begin{bmatrix} a^3_{11} & a^3_{12} & a^3_{13} \\ a^3_{21} & a^3_{22} & a^3_{23} \\ a^3_{31} & a^3_{32} & a^3_{33} \end{bmatrix}$.

It is helpful to define,

$$\xi_t = \begin{bmatrix} \Delta y_t \\ \Delta y_{t-1} \\ \alpha' y_t \end{bmatrix}, F = \begin{bmatrix} -(A_2+A_3) & -A_3 & \gamma \\ I_3 & 0 & 0 \\ -\alpha'(A_2+A_3) & -\alpha' A_3 & \alpha'\gamma+1 \end{bmatrix} \text{ and } \nu_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \alpha'\varepsilon_t \end{bmatrix},$$

to arrive at,

$$\xi_t = F\xi_{t-1} + \nu_t. \tag{24}$$

If we consider cointegration and common-cycle restrictions, the following relations hold:

(i) $\widetilde{\alpha}' A_3 = 0 \Rightarrow A_3 = \begin{bmatrix} Ga^3_{31} & Ga^3_{32} & Ga^3_{33} \\ Ka^3_{31} & Ka^3_{32} & Ka^3_{33} \\ a^3_{31} & a^3_{32} & a^3_{33} \end{bmatrix}$, where $G = [R_{21}K + R_{31}]$, $K = (R_{32} - R_{31})/(R_{21} - R_{22})$, $R_{i1} = \widetilde{\alpha}_{i1}/\widetilde{\alpha}_{11}$ and $R_{i2} = \widetilde{\alpha}_{i2}/\widetilde{\alpha}_{12}$ $(i = 2, 3)$,

(ii) $\widetilde{\alpha}'(A_2 + A_3) = 0 \Rightarrow \widetilde{\alpha}' A_2 = 0 \Rightarrow A_2 = \begin{bmatrix} Ga^2_{31} & Ga^2_{32} & Ga^2_{33} \\ Ka^2_{31} & Ka^2_{32} & Ka^2_{33} \\ a^2_{31} & a^2_{32} & a^2_{33} \end{bmatrix}$,

(iii) $\widetilde{\alpha}'\gamma = 0 \Rightarrow \gamma = \begin{bmatrix} G\gamma_{31} \\ K\gamma_{31} \\ \gamma_{31} \end{bmatrix}$,

(iv) $\alpha'(A_2 + A_3) = \begin{bmatrix} (a^2_{31} + a^3_{31})S & (a^2_{32} + a^3_{32})S & (a^2_{33} + a^3_{33})S \end{bmatrix}$ and $\alpha' A_3 = \begin{bmatrix} a^3_{31}S & a^3_{32}S & a^3_{33}S \end{bmatrix}$, where $S = \alpha_{11}G + \alpha_{21}K + \alpha_{31}$,

(v) $\alpha'\gamma + 1 = \gamma_{31}S + 1$.

The restrictions above imply that:

$$F = \begin{bmatrix} -G(a^2_{31}+a^3_{31}) & -G(a^2_{32}+a^3_{32}) & -G(a^2_{33}+a^3_{33}) & -Ga^3_{31} & -Ga^3_{32} & -Ga^3_{33} & G\gamma_{31} \\ -K(a^2_{31}+a^3_{31}) & -K(a^2_{32}+a^3_{32}) & -K(a^2_{33}+a^3_{33}) & -Ka^3_{31} & -Ka^3_{32} & -Ka^3_{33} & K\gamma_{31} \\ -(a^2_{31}+a^3_{31}) & -(a^2_{32}+a^3_{32}) & -(a^2_{33}+a^3_{33}) & -a^3_{31} & -a^3_{32} & -a^3_{33} & \gamma_{31} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -(a^2_{31}+a^3_{31})S & -(a^2_{32}+a^3_{32})S & -(a^2_{33}+a^3_{33})S & -a^3_{31}S & -a^3_{32}S & -a^3_{33}S & \gamma_{31}S+1 \end{bmatrix}.$$

If all the eigenvalues of matrix F lie inside the unit circle, then the VAR (24) is covariance-stationary. The eigenvalue of the matrix F is a number λ such that

$$|F - \lambda I_7| = 0. \tag{25}$$

The solution of (25) is:

$$\begin{aligned} 0 \;=\; & \lambda^7 - \left[(1+\gamma_{31}S) - \left(a_{33}^3 + a_{32}^3 K + a_{31}^3 G\right) - \left(a_{33}^2 + a_{32}^2 K + a_{31}^2 G\right)\right]\lambda^6 \\ & - \left(a_{33}^2 + a_{32}^2 K + a_{31}^2 G\right)\lambda^5 - \left(a_{33}^3 + a_{32}^3 K + a_{31}^3 G\right)\lambda^4. \end{aligned} \tag{26}$$

If we define $\Omega = -\left[(1+\gamma_{31}S) - \left(a_{33}^3 + a_{32}^3 K + a_{31}^3 G\right) - \left(a_{33}^2 + a_{32}^2 K + a_{31}^2 G\right)\right]$, $\Theta = -\left(a_{33}^2 + a_{32}^2 K + a_{31}^2 G\right)$ and $\Psi = -\left(a_{33}^3 + a_{32}^3 K + a_{31}^3 G\right)$, (26) is:

$$\lambda^7 + \Omega\lambda^6 + \Theta\lambda^5 + \Psi\lambda^4 = 0. \tag{27}$$

The roots of this polynomial are $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$, $\lambda_5 = A + B - \frac{\Omega}{3}$, $\lambda_6 = A\omega + B\omega^2 - \frac{\Omega}{3}$, $\lambda_7 = A\omega^2 + B\omega - \frac{\Omega}{3}$, where, $\omega = \frac{-1+\sqrt[2]{3}}{2}$, $A = \sqrt[3]{-\frac{b^2}{2} + \sqrt[2]{\frac{b^2}{4} + \frac{a^3}{27}}}$, $B = \sqrt[3]{-\frac{b^2}{2} - \sqrt[2]{\frac{b^2}{4} + \frac{a^3}{27}}}$, $a = \frac{1}{3}\left(3\Theta - \Omega^2\right)$ and $b = \frac{1}{27}\left(2\Omega^3 - 9\Omega\Theta + 27\Psi\right)$.

Using these restrictions we can guarantee that cointegration and SCCF restrictions hold for well behaved VECMs being simulated.

B Tables

Table 1

Percentage improvement in different forecast accuracy measures when the true restrictions (lags, rank and number of cointegrated vectors) are imposed and the true models are trivariate (3,1,1).

Horizon	VECM over VAR in levels			Reduced-rank VECM over the VAR in levels			Reduced-rank VECM over the VECM		
(h)	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE
				Low system R^2 Measure					
Sample size 100									
1	5.11	5.11	1.67	16.40	16.40	5.34	11.28	11.28	3.68
4	20.48	407.50	99.96	44.55	409.99	102.33	24.06	2.50	2.37
8	33.39	635.26	171.01	60.44	635.99	172.38	27.06	0.73	1.37
12	44.08	769.22	212.98	72.74	769.63	213.95	28.66	0.41	0.97
16	54.73	866.74	242.75	84.61	867.05	243.51	29.89	0.31	0.76
Sample size 200									
1	1.92	1.92	0.62	7.30	7.30	2.38	5.38	5.38	1.76
4	7.50	381.37	93.96	17.99	382.16	94.99	10.49	0.79	1.03
8	12.66	594.71	161.02	24.21	594.91	161.59	11.56	0.20	0.57
12	16.12	719.71	200.69	28.26	719.84	201.08	12.14	0.13	0.40
16	19.20	809.72	228.77	31.71	809.75	229.07	12.51	0.03	0.30
				High system R^2 Measure					
Sample size 100									
1	-165.41	-165.41	-85.26	-30.33	-30.33	-13.73	135.08	135.08	71.53
4	-278.95	275.06	49.22	-197.63	273.60	64.62	81.32	-1.46	15.40
8	-321.63	442.14	103.05	-288.41	443.31	110.86	33.22	1.17	7.81
12	-330.03	537.34	138.04	-323.14	537.74	143.33	6.89	0.39	5.29
16	-325.15	607.36	165.11	-330.58	607.47	169.10	-5.43	0.11	3.98
Sample size 200									
1	-169.09	-169.09	-88.58	-40.54	-40.54	-16.87	128.55	128.55	71.72
4	-287.72	261.16	45.16	-230.40	256.60	58.97	57.33	-4.56	13.80
8	-336.24	421.61	97.18	-334.98	421.94	103.68	1.25	0.33	6.50
12	-350.81	512.31	130.20	-379.24	512.47	134.52	-28.43	0.15	4.33
16	-352.76	577.82	155.52	-395.17	577.84	158.76	-42.41	0.02	3.24

GFESM is Clements and Hendry's generalized forecast error second moment measure, |MSFE| is the determinant of the of the mean squared forecast error matrix and TMSFE is the trace of the MSFE matrix.

Table 2 - Low system R^2 Measure

Frequency of lag (p) and cointegrating vectors (q) choice by different criteria when the true model is (3,1,1) in levels.

		Number of Observations = 100				Number of Observations = 200			
		Selected Cointegrating Vectors				Selected Cointegrating Vectors			
	Selected Lag	0	**1**	2	3	0	**1**	2	3
AIC(p)	1	0.01	0.36	0.07	0.05	0.00	0.00	0.00	0.00
	2	35.79	3.20	0.45	0.47	7.66	1.85	0.44	0.51
	3	40.32	**8.23**	1.06	0.60	42.60	**31.47**	5.14	3.98
	4	4.13	1.42	0.20	0.08	2.46	1.96	0.35	0.24
	5	1.14	0.56	0.07	0.02	0.41	0.41	0.07	0.04
	6	0.46	0.32	0.06	0.02	0.11	0.13	0.03	0.01
	7	0.27	0.21	0.04	0.01	0.04	0.04	0.00	0.01
	8	0.16	0.20	0.04	0.01	0.01	0.03	0.01	0.00
HQ(p)	1	0.26	2.53	0.50	0.24	0.00	0.00	0.00	0.00
	2	68.50	5.66	0.78	0.87	35.07	8.10	1.86	2.39
	3	15.34	**4.26**	0.56	0.27	24.47	**21.95**	3.46	2.63
	4	0.14	0.07	0.02	0.00	0.03	0.03	0.01	0.00
	5	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SC(p)	1	3.08	11.69	2.10	1.13	0.02	0.05	0.02	0.00
	2	72.73	5.59	0.75	0.88	66.71	15.02	3.38	4.48
	3	1.34	**0.63**	0.07	0.03	3.90	**5.12**	0.78	0.53
	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Numbers represent the percentage times that the model selection criterion chose that cell, corresponding to the lag and number of cointegrating vectors, in 100,000 times (1000 simulations of 100 different DGPs). The true lag-cointegrating vectors are indentified by bold numbers.

Table 3 - High system R^2 Measure

Frequency of lag (p) and cointegrating vectors (q) choice by different criteria when the true model is (3,1,1) in levels.

		Number of Observations = 100				Number of Observations = 200			
		Selected Cointegrating Vectors				Selected Cointegrating Vectors			
	Selected Lag	0	**1**	2	3	0	**1**	2	3
AIC(p)	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.00	32.93	2.69	1.37	0.00	8.64	1.08	0.67
	3	0.00	**49.10**	3.50	1.43	0.00	**71.47**	7.57	4.47
	4	0.01	5.24	0.46	0.16	0.00	4.17	0.47	0.26
	5	0.03	1.42	0.16	0.06	0.00	0.73	0.08	0.06
	6	0.04	0.59	0.08	0.02	0.00	0.20	0.03	0.02
	7	0.04	0.30	0.05	0.01	0.00	0.06	0.01	0.00
	8	0.04	0.23	0.05	0.01	0.00	0.02	0.00	0.00
HQ(p)	1	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00
	2	0.00	68.34	5.55	2.87	0.00	39.86	5.15	3.17
	3	0.00	**20.69**	1.60	0.66	0.00	**44.22**	4.73	2.82
	4	0.00	0.23	0.02	0.01	0.00	0.05	0.01	0.00
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SC(p)	1	0.00	0.42	0.10	0.06	0.00	0.00	0.00	0.00
	2	0.00	86.26	6.96	3.60	0.00	74.78	9.46	6.00
	3	0.00	**2.35**	0.17	0.08	0.00	**8.31**	0.95	0.50
	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Numbers represent the percentage times that the model selection criterion chose that cell, corresponding to the lag and number of cointegrating vectors, in 100,000 times (1000 simulations of 100 different DGPs). The true lag-cointegrating vectors are indentified by bold numbers.

Table 4 - Low system R^2 Measure

Frequency of lag-rank (p,r) and cointegrating vectors (q) choice by different criteria when the true model is (2,1,1) in differences.

		Number of Observations = 100												Number of Observations = 200											
	Selected Number of coint. vectors	0			**1**			2			3			0			**1**			2			3		
	Selected Rank	1	2	3	**1**	2	3	1	2	3	1	2	3	1	2	3	**1**	2	3	1	2	3	1	2	3
	Selected Lag																								
AIC(p,r)	1	8.69	0.19	0.00	0.82	0.03	0.00	0.11	0.00	0.00	0.12	0.00	0.00	0.76	0.02	0.00	0.20	0.01	0.00	0.05	0.00	0.00	0.07	0.00	0.00
	2	48.48	1.08	0.01	**8.06**	0.29	0.00	1.01	0.05	0.00	0.62	0.02	0.00	40.25	0.63	0.00	**28.07**	0.55	0.00	4.55	0.10	0.00	3.60	0.08	0.00
	3	8.83	0.42	0.00	2.16	0.13	0.00	0.24	0.03	0.00	0.09	0.01	0.00	5.71	0.21	0.00	4.15	0.15	0.00	0.60	0.04	0.00	0.40	0.01	0.00
	4	3.83	0.20	0.00	1.28	0.10	0.00	0.14	0.01	0.00	0.05	0.00	0.00	2.24	0.06	0.00	1.71	0.07	0.00	0.24	0.01	0.00	0.13	0.00	0.00
	5	2.42	0.12	0.00	1.07	0.08	0.00	0.13	0.01	0.00	0.03	0.00	0.00	1.15	0.03	0.00	0.92	0.04	0.00	0.12	0.00	0.00	0.06	0.01	0.00
	6	1.74	0.09	0.00	0.88	0.08	0.00	0.11	0.01	0.00	0.03	0.00	0.00	0.69	0.01	0.00	0.56	0.01	0.00	0.06	0.00	0.00	0.03	0.00	0.00
	7	1.45	0.08	0.00	1.01	0.08	0.00	0.14	0.01	0.00	0.03	0.00	0.00	0.42	0.01	0.00	0.40	0.01	0.00	0.04	0.00	0.00	0.02	0.00	0.00
	8	1.58	0.07	0.00	1.23	0.10	0.00	0.21	0.02	0.00	0.05	0.00	0.00	0.35	0.01	0.00	0.31	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00
HQ(p,r)	1	23.50	0.02	0.00	2.17	0.01	0.00	0.30	0.00	0.00	0.33	0.00	0.00	4.03	0.00	0.00	1.03	0.00	0.00	0.22	0.00	0.00	0.30	0.00	0.00
	2	55.15	0.04	0.00	**9.88**	0.02	0.00	1.25	0.00	0.00	0.71	0.00	0.00	47.49	0.01	0.00	**33.93**	0.01	0.00	5.54	0.00	0.00	4.29	0.00	0.00
	3	3.45	0.01	0.00	1.10	0.00	0.00	0.12	0.00	0.00	0.05	0.00	0.00	1.33	0.00	0.00	1.13	0.00	0.00	0.16	0.00	0.00	0.10	0.00	0.00
	4	0.72	0.00	0.00	0.35	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.17	0.00	0.00	0.15	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00
	5	0.25	0.00	0.00	0.17	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	6	0.04	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.02	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SC(p,r)	1	45.68	0.00	0.00	4.08	0.00	0.00	0.56	0.00	0.00	0.60	0.00	0.00	14.55	0.00	0.00	3.48	0.00	0.00	0.77	0.00	0.00	1.00	0.00	0.00
	2	38.94	0.00	0.00	**7.87**	0.00	0.00	0.97	0.00	0.00	0.54	0.00	0.00	40.30	0.00	0.00	**30.87**	0.00	0.00	4.97	0.00	0.00	3.83	0.00	0.00
	3	0.46	0.00	0.00	0.20	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.11	0.00	0.00	0.10	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	4	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Numbers represent the percentage times that the model selection criterion chose that cell, corresponding to the lag-rank and number of cointegrating vectors, in 100,00
(1000 simulations of 100 different DGPs). The true lag-rank-cointegrating vectors are indentified by bold numbers.

Table 5 - High system R^2 Measure

Frequency of lag-rank (p,r) and cointegrating vectors (q) choice by different criteria when the true model is (2,1,1) in differences.

		Number of Observations = 100												Number of Observations = 200											
	Selected Number of coint. vectors	0			**1**			2			3			0			**1**			2			3		
	Selected Rank	1	2	3	**1**	2	3	1	2	3	1	2	3	1	2	3	**1**	2	3	1	2	3	1	2	3
	Selected Lag																								
AIC(p,r)	1	0.00	0.00	0.00	8.90	0.18	0.00	0.73	0.03	0.00	0.35	0.01	0.00	0.00	0.00	0.00	0.95	0.02	0.00	0.12	0.00	0.00	0.07	0.00	0.00
	2	0.00	0.00	0.00	**54.82**	1.27	0.01	3.62	0.15	0.00	1.48	0.05	0.00	0.00	0.00	0.00	**66.01**	1.15	0.00	6.87	0.14	0.00	4.08	0.07	0.00
	3	0.01	0.00	0.00	10.27	0.49	0.00	0.65	0.06	0.00	0.23	0.02	0.00	0.00	0.00	0.00	9.30	0.30	0.00	0.92	0.04	0.00	0.51	0.03	0.00
	4	0.18	0.01	0.00	4.51	0.27	0.00	0.33	0.04	0.00	0.07	0.01	0.00	0.00	0.00	0.00	3.70	0.13	0.00	0.33	0.01	0.00	0.17	0.01	0.00
	5	0.31	0.01	0.00	2.77	0.16	0.00	0.17	0.02	0.00	0.05	0.01	0.00	0.00	0.00	0.00	1.97	0.07	0.00	0.16	0.01	0.00	0.08	0.00	0.00
	6	0.40	0.02	0.00	1.88	0.11	0.00	0.14	0.01	0.00	0.04	0.00	0.00	0.00	0.00	0.00	1.11	0.02	0.00	0.10	0.01	0.00	0.04	0.00	0.00
	7	0.46	0.03	0.00	1.61	0.11	0.00	0.13	0.02	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.76	0.01	0.00	0.07	0.00	0.00	0.03	0.00	0.00
	8	0.62	0.05	0.00	1.77	0.10	0.00	0.16	0.03	0.00	0.03	0.00	0.00	0.02	0.00	0.00	0.54	0.01	0.00	0.04	0.00	0.00	0.01	0.00	0.00
HQ(p,r)	1	0.00	0.00	0.00	23.56	0.03	0.00	1.97	0.01	0.00	0.99	0.00	0.00	0.00	0.00	0.00	5.00	0.00	0.00	0.63	0.00	0.00	0.38	0.00	0.00
	2	0.00	0.00	0.00	**61.57**	0.05	0.00	4.10	0.01	0.00	1.70	0.00	0.00	0.00	0.00	0.00	**78.00**	0.01	0.00	8.16	0.00	0.00	4.85	0.00	0.00
	3	0.01	0.00	0.00	4.07	0.01	0.00	0.27	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	2.26	0.00	0.00	0.22	0.00	0.00	0.11	0.00	0.00
	4	0.04	0.00	0.00	0.87	0.00	0.00	0.06	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00
	5	0.03	0.00	0.00	0.28	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	6	0.01	0.00	0.00	0.05	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SC(p,r)	1	0.00	0.00	0.00	46.03	0.00	0.00	3.74	0.00	0.00	1.94	0.00	0.00	0.00	0.00	0.00	17.37	0.00	0.00	2.24	0.00	0.00	1.38	0.00	0.00
	2	0.00	0.00	0.00	**43.48**	0.00	0.00	2.90	0.00	0.00	1.21	0.00	0.00	0.00	0.00	0.00	**67.60**	0.00	0.00	7.03	0.00	0.00	4.19	0.00	0.00
	3	0.01	0.00	0.00	0.58	0.00	0.00	0.04	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00
	4	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Numbers represent the percentage times that the model selection criterion chose that cell, corresponding to the lag-rank and number of cointegrating vectors, in 100,00
(1000 simulations of 100 different DGPs). The true lag-rank-cointegrating vectors are indentified by bold numbers.

Table 6 - Low system R^2 Measure

Percentage improvement in different measures of accuracy in forecast generated by the possibly reduced-rank VECM over the VECM chosen by the same model selection criterion when the true models are trivariate (3,1,1).

Horizon	AIC			HQ			SC		
(h)	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE
Sample size 100									
1	7.13	7.13	2.19	10.55	10.55	**3.59**	19.75	19.75	7.27
4	7.44	0.46	0.99	14.09	0.22	1.19	26.87	0.74	**2.06**
8	4.01	0.00	0.43	14.62	0.16	0.61	30.12	0.77	**1.15**
12	3.28	-0.12	0.28	15.01	0.05	0.41	32.42	0.57	**0.83**
16	2.95	-0.10	0.21	15.43	0.14	0.32	34.57	0.57	**0.68**
Sample size 200									
1	5.01	5.01	1.62	6.63	6.63	**2.23**	7.85	7.85	2.73
4	8.30	0.60	0.87	11.10	0.42	**0.97**	12.53	0.03	0.93
8	8.12	0.15	0.46	12.07	0.13	**0.51**	14.05	0.17	0.49
12	8.47	0.04	0.32	12.69	0.05	**0.36**	15.04	0.05	0.35
16	8.75	0.03	0.24	13.13	0.03	**0.28**	15.70	0.01	0.27

GFESM is Clements and Hendry's generalized forecast error second moment measure, |MSFE| is the determinant of the of the mean squared forecast error matrix and TMSFE is the trace of the MSFE matrix. Bold and underline numbers denote, respectively, the best and the worst forecasting performance across all three information criteria based on TMSFE.

Table 7 - High system R^2 Measure

Percentage improvement in different measures of accuracy in forecast generated by the possibly reduced-rank VECM over the VECM chosen by the same model selection criterion when the true models are trivariate (3,1,1).

Horizon	AIC			HQ			SC		
(h)	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE
Sample size 100									
1	138.40	138.40	**71.77**	110.49	110.49	58.76	95.76	95.76	51.07
4	103.73	1.73	**17.20**	55.33	-3.22	11.24	34.27	-3.56	8.43
8	56.88	1.11	**8.77**	9.19	0.54	5.46	-7.61	0.35	4.01
12	27.76	0.09	**5.90**	-16.36	0.17	3.66	-28.33	0.10	2.67
16	10.11	-0.06	**4.43**	-29.33	0.03	2.74	-37.80	0.03	2.00
Sample size 200									
1	133.34	133.34	**73.70**	117.91	117.91	65.43	99.75	99.75	54.73
4	76.12	-2.42	**15.59**	52.55	-5.22	12.20	36.39	-6.38	9.24
8	20.25	0.17	**7.44**	-1.39	0.18	5.61	-14.53	-0.02	4.08
12	-11.63	0.06	**4.92**	-30.73	0.16	3.72	-41.80	0.06	2.68
16	-29.18	-0.03	**3.68**	-44.92	0.00	2.79	-54.58	0.01	2.00

GFESM is Clements and Hendry's generalized forecast error second moment measure, |MSFE| is the determinant of the of the mean squared forecast error matrix and TMSFE is the trace of the MSFE matrix. Bold and underline numbers denote, respectively, the best and the worst forecasting performance across all three information criteria based on TMSFE.

Table 8 - Low system R^2 Measure

Percentage improvement in different measures of accuracy in forecast generated by the VECM over the VAR in levels when the true models are trivariate (3,1,1).

Horizon	AIC			HQ			SC		
(h)	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE
Sample size 100									
1	10.31	10.31	3.36	10.98	10.98	**3.54**	10.47	10.47	3.28
4	36.52	413.09	101.68	38.03	413.81	**101.86**	39.19	420.44	102.98
8	57.11	640.17	172.27	56.50	638.34	**171.44**	57.17	645.17	172.04
12	72.72	774.16	214.09	70.62	771.59	**212.73**	70.63	779.15	213.20
16	87.01	871.35	243.82	83.48	868.17	**242.18**	82.35	875.86	242.63
Sample size 200									
1	1.71	1.71	**0.53**	2.53	2.53	0.78	3.47	3.47	1.07
4	6.38	381.87	94.00	7.80	382.43	**95.05**	9.21	383.72	94.27
8	11.09	595.67	161.08	11.98	595.64	**161.35**	12.64	596.16	160.59
12	14.40	720.74	200.77	14.94	720.74	**200.70**	15.01	721.30	199.88
16	17.27	810.68	228.87	17.55	810.76	**228.63**	17.21	811.28	227.77

GFESM is Clements and Hendry's generalized forecast error second moment measure, |MSFE| is the determinant of the of the mean squared forecast error matrix and TMSFE is the trace of the MSFE matrix. Bold and underline numbers denote, respectively, the best and the worst forecasting performance across all three information criteria based on TMSFE.

Table 9 - High system R^2 Measure

Percentage improvement in different measures of accuracy in forecast generated by the VECM over the VAR in levels when the true models are trivariate (3,1,1).

Horizon	AIC			HQ			SC		
(h)	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE
Sample size 100									
1	-162.81	-162.81	-82.96	-139.92	-139.92	-72.04	-126.87	-126.87	**-65.06**
4	-285.21	273.12	48.54	-264.51	274.67	51.43	-254.78	275.11	**52.73**
8	-331.16	441.54	102.48	-310.61	439.34	103.83	-302.50	438.34	**104.37**
12	-339.86	537.27	137.61	-319.49	534.17	138.14	-312.58	532.86	**138.32**
16	-334.70	607.62	164.82	-314.97	604.05	164.83	-309.10	602.53	**164.78**
Sample size 200									
1	-169.58	-169.58	-88.65	-155.64	-155.64	-81.24	-135.48	-135.48	**-69.89**
4	-290.20	260.25	44.80	-281.78	261.19	**100.02**	-269.76	262.29	48.58
8	-339.60	421.49	96.95	-333.23	420.79	**148.56**	-323.40	420.30	98.72
12	-354.77	512.23	130.02	-348.81	511.35	**180.21**	-339.44	510.66	130.99
16	-357.24	577.86	155.40	-351.48	576.90	**204.83**	-342.52	576.13	155.92

GFESM is Clements and Hendry's generalized forecast error second moment measure, |MSFE| is the determinant of the of the mean squared forecast error matrix and TMSFE is the trace of the MSFE matrix. Bold and underline numbers denote, respectively, the best and the worst forecasting performance across all three information criteria based on TMSFE.

Table 10 - Low system R^2 Measure

Percentage improvement in different measures of accuracy in forecast generated by the possibly reduced-rank VECM over the VAR in levels when the true models are trivariate (3,1,1).

Horizon	AIC			HQ			SC		
(h)	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE
Sample size 100									
1	17.44	17.44	5.55	21.53	21.53	**7.13**	30.22	30.22	10.55
4	43.97	413.55	102.68	52.12	414.03	103.06	66.06	421.18	**105.04**
8	61.11	640.17	172.70	71.13	638.51	172.05	87.30	645.94	**173.19**
12	76.00	774.04	214.37	85.63	771.64	213.15	103.04	779.73	**214.04**
16	89.96	871.25	244.03	98.91	868.31	242.50	116.92	876.43	**243.30**
Sample size 200									
1	6.72	6.72	2.15	9.16	9.16	**3.00**	11.32	11.32	3.80
4	14.68	382.47	94.87	18.90	382.85	**95.07**	21.74	383.75	95.19
8	19.21	595.82	161.54	24.05	595.77	**161.36**	26.69	596.33	161.07
12	22.87	720.77	201.09	27.64	720.79	**200.71**	30.05	721.35	200.23
16	26.02	810.71	229.11	30.68	810.79	**228.64**	32.91	811.29	228.04

GFESM is Clements and Hendry's generalized forecast error second moment measure, |MSFE| is the determinant of the of the mean squared forecast error matrix and TMSFE is the trace of the MSFE matrix. Bold and underline numbers denote, respectively, the best and the worst forecasting performance across all three information criteria based on TMSFE.

Table 11 - High system R^2 Measure

Percentage improvement in different measures of accuracy in forecast generated by the possibly reduced-rank VECM over the VAR in levels when the true models are trivariate (3,1,1).

Horizon	AIC			HQ			SC		
(h)	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE	GFESM	|MSFE|	TMSFE
Sample size 100									
1	-24.40	-24.40	**-11.19**	-29.44	-29.44	-13.27	-31.11	-31.11	-14.00
4	-181.47	274.86	**65.75**	-209.18	271.45	62.68	-220.52	271.55	61.16
8	-274.28	442.65	**111.26**	-301.42	439.88	109.29	-310.11	438.69	108.38
12	-312.10	537.37	**143.51**	-335.85	534.34	141.80	-340.91	532.96	140.99
16	-324.59	607.56	**169.24**	-344.29	604.08	167.58	-346.90	602.57	166.78
Sample size 200									
1	-36.25	-36.25	**-14.95**	-37.73	-37.73	-15.80	-35.74	-35.74	-15.16
4	-214.07	257.82	**60.39**	-229.23	255.97	58.65	-233.38	255.91	57.82
8	-319.35	421.66	**104.38**	-334.61	420.97	103.35	-337.93	420.29	102.80
12	-366.40	512.29	**134.95**	-379.55	511.51	134.16	-381.24	510.72	133.67
16	-386.42	577.83	**159.08**	-396.40	576.90	158.39	-397.09	576.14	157.92

GFESM is Clements and Hendry's generalized forecast error second moment measure, |MSFE| is the determinant of the of the mean squared forecast error matrix and TMSFE is the trace of the MSFE matrix. Bold and underline numbers denote, respectively, the best and the worst forecasting performance across all three information criteria based on TMSFE.

Table 12a: Percentage Improvement on MSE						
Restricted VECM over VECM						
$n = 3$, $q = 1$, $r = 1$						
	Systems with high R^2					
	Long-Term Coeffs.		Short-Term Coeffs.		All Coeffs.	
N	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$
100	118.24	14.79	1917.04	56.80	1751.00	47.25
200	113.98	15.77	2286.29	61.61	2108.90	57.77
Individual Results: Long-Term Coefficients						
N	γ_1	γ_2	γ_3	α_1	α_2	
100	21.97	29.40	36.91	6.20	7.40	
200	23.97	29.45	39.60	3.51	3.44	
Individual Results: Short-Term Coefficients						
N	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}
100	42.99	40.38	57.81	67.19	70.38	76.86
200	44.54	47.31	54.49	72.67	70.30	83.60
Individual Results: Short-Term Coefficients						
N	ϕ_{31}	ϕ_{32}	ϕ_{33}			
100	99.12	102.82	110.26			
200	103.23	113.29	120.17			

Notes: $|MSE|$ denotes the determinant of the mean-squared error matrix, while $TMSE$ denotes its trace. Parameters in the restricted VECM have the following labels:

$$\Delta y_t = \underset{(3\times1)}{\boldsymbol{\gamma}}\underset{(1\times3)}{\boldsymbol{\alpha}}' y_{t-1} - \underset{(3\times1)}{\boldsymbol{\lambda}}\underset{(1\times3)}{\boldsymbol{\delta}'} \Delta y_{t-1} + \varepsilon_t \quad (28)$$

$$= \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \begin{bmatrix} 1 & \alpha_1 & \alpha_2 \end{bmatrix} y_{t-1} - \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \Delta y_{t-1} + \boldsymbol{\varepsilon}_t \quad (29)$$

Table 12b: Percentage Improvement on MSE						
Restricted VECM over VECM						
$n = 3$, $q = 1$, $r = 1$						
	Systems with low R^2					
	Long-Term Coeffs.		Short-Term Coeffs.		All Coeffs.	
N	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$
100	70.45	8.47	1396.64	39.53	1339.71	20.92
200	51.18	9.19	1785.18	44.73	1714.05	40.53
Individual Results: Long-Term Coefficients						
N	γ_1	γ_2	γ_3	α_1	α_2	
100	14.70	12.89	15.84	-6.88	9.07	
200	13.90	11.20	14.90	4.86	5.38	
Individual Results: Short-Term Coefficients						
N	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}
100	52.10	44.12	52.15	52.44	50.37	46.71
200	53.99	51.15	48.81	61.17	50.34	54.35
Individual Results: Short-Term Coefficients						
N	ϕ_{31}	ϕ_{32}	ϕ_{33}			
100	70.37	66.85	63.10			
200	75.73	79.18	73.43			

Notes: Same as in Table 12a.

Table 13a: Percentage Improvement on MSE						
Restricted VECM over VECM						
$n = 3$, $q = 1$, $r = 2$						
	Systems with high R^2					
	Long-Term Coeffs.		Short-Term Coeffs.		All Coeffs.	
N	$\lvert MSE \rvert$	$TMSE$	$\lvert MSE \rvert$	$TMSE$	$\lvert MSE \rvert$	$TMSE$
100	21.97	8.49	147.75	7.68	149.56	13.39
200	15.44	2.83	254.94	10.30	237.53	9.96
Individual Results: Long-Term Coefficients						
N	γ_1	γ_2	γ_3	α_1	α_2	
100	4.28	4.66	2.99	7.77	5.45	
200	5.58	4.97	3.04	0.62	0.41	
Individual Results: Short-Term Coefficients						
N	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}
100	18.19	14.46	-0.95	22.17	18.23	-0.43
200	24.22	19.84	-0.70	28.00	22.50	-0.38
Individual Results: Short-Term Coefficients						
N	ϕ_{31}	ϕ_{32}	ϕ_{33}			
100	15.25	12.58	-4.73			
200	17.25	16.56	-2.30			

Notes: $|MSE|$ denotes the determinant of the mean-squared error matrix, while $TMSE$ denotes its trace. Parameters in the restricted VECM have the following labels:

$$\Delta y_t = \underset{(3\times1)}{\boldsymbol{\gamma}}\underset{(1\times3)}{\boldsymbol{\alpha}}{}' y_{t-1} - \underset{(3\times2)}{\boldsymbol{\lambda}}\underset{(2\times3)}{\boldsymbol{\delta}'} \Delta y_{t-1} + \varepsilon_t \tag{30}$$

$$= \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \begin{bmatrix} 1 & \alpha_1 & \alpha_2 \end{bmatrix} y_{t-1} - \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \Delta y_{t-1} + \boldsymbol{\varepsilon}_t \tag{31}$$

Table 13b: Percentage Improvement on MSE						
Restricted VECM over VECM						
$n = 3$, $q = 1$, $r = 2$						
	Systems with low R^2					
	Long-Term Coeffs.		Short-Term Coeffs.		All Coeffs.	
N	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$
100	34.64	6.46	161.56	6.63	181.13	7.03
200	16.07	2.74	292.45	9.70	282.62	8.08
Individual Results: Long-Term Coefficients						
N	γ_1	γ_2	γ_3	α_1	α_2	
100	4.00	0.64	6.38	14.64	1.69	
200	4.66	1.96	6.25	1.79	1.28	
Individual Results: Short-Term Coefficients						
N	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}
100	10.78	9.15	-2.58	7.50	4.87	-1.28
200	14.53	14.76	-0.02	11.24	8.59	-0.45
Individual Results: Short-Term Coefficients						
N	ϕ_{31}	ϕ_{32}	ϕ_{33}			
100	28.85	24.63	-0.03			
200	33.35	35.30	0.33			

Notes: Same as Table 13a.

Table 14a: Percentage Improvement on MSE						
Restricted VECM over VECM						
$n=3$, $q=2$, $r=2$						
	Systems with high R^2					
	Long-Term Coeffs.		Short-Term Coeffs.		All Coeffs.	
N	$\lvert MSE\rvert$	$TMSE$	$\lvert MSE\rvert$	$TMSE$	$\lvert MSE\rvert$	$TMSE$
100	31.52	19.79	656.35	303.04	709.41	222.43
200	25.55	10.00	834.81	381.13	875.44	307.15
Individual Results: Long-Term Coefficients						
N	γ_{11}^E	γ_{21}^E	γ_{31}^E	γ_{12}^E	γ_{22}^E	γ_{32}^E
100	3.35	2.68	2.27	1.32	2.86	2.21
200	3.93	3.82	2.88	2.10	3.72	1.97
Individual Results: Long-Term Coefficients						
N	α_{21}^E	α_{22}^E				
100	16.70	17.89				
200	7.81	7.83				
Individual Results: Short-Term Coefficients						
N	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}
100	212.81	284.27	154.87	259.75	344.40	63.81
200	274.24	360.95	218.85	332.52	426.55	106.48
Individual Results: Short-Term Coefficients						
N	ϕ_{31}	ϕ_{32}	ϕ_{33}			
100	175.70	124.21	199.58			
200	233.47	172.88	257.02			

Notes: $|MSE|$ denotes the determinant of the mean-squared error matrix, while $TMSE$ denotes its trace. Parameters in the restricted VECM, put in canonical echelon form, have the following labels:

$$\Delta y_t = \underset{(3\times 2)}{\boldsymbol{\gamma}}\underset{(2\times 3)}{\boldsymbol{\alpha}}' y_{t-1} - \underset{(3\times 2)}{\boldsymbol{\lambda}}\underset{(2\times 3)}{\boldsymbol{\delta}'}\Delta y_{t-1} + \boldsymbol{\varepsilon}_t \tag{32}$$

$$= \begin{bmatrix} \gamma_{11}^E & \gamma_{12}^E \\ \gamma_{21}^E & \gamma_{22}^E \\ \gamma_{31}^E & \gamma_{32}^E \end{bmatrix} \begin{bmatrix} 1 & 0 & \alpha_{21}^E \\ 0 & 1 & \alpha_{22}^E \end{bmatrix} y_{t-1} - \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \Delta y_{t-1} \tag{33}$$

$$+\boldsymbol{\varepsilon}_t \tag{34}$$

Table 14b: Percentage Improvement on MSE						
Restricted VECM over VECM						
$n = 3$, $q = 2$, $r = 2$						
	Systems with low R^2					
	Long-Term Coeffs.		Short-Term Coeffs.		All Coeffs.	
N	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$	$\|MSE\|$	$TMSE$
100	85.91	20.55	1164.25	218.65	1232.60	94.97
200	67.39	13.58	1588.65	287.59	1632.52	209.33
Individual Results: Long-Term Coefficients						
N	γ_{11}^E	γ_{21}^E	γ_{31}^E	γ_{12}^E	γ_{22}^E	γ_{32}^E
100	7.29	3.83	4.98	13.39	10.43	10.02
200	6.65	3.25	3.92	13.50	10.81	11.92
Individual Results: Long-Term Coefficients						
N	α_{21}^E	α_{22}^E				
100	26.61	18.43				
200	10.15	8.77				
Individual Results: Short-Term Coefficients						
N	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}
100	150.02	142.22	166.05	158.09	178.03	174.47
200	199.59	206.38	233.70	205.15	244.52	229.52
Individual Results: Short-Term Coefficients						
N	ϕ_{31}	ϕ_{32}	ϕ_{33}			
100	164.28	167.71	177.95			
200	224.57	232.45	236.73			

Notes: Same as Table 14a.