FUNDAÇÃO
GETULIO VARGAS

EPGE
Escola de Pós-Graduação
em Economia

Ensaios Econômicos

Escola de
Pós-Graduação
em Economia
da Fundação
Getulio Vargas

N° 352 ISSN 0104-8910

The Importance of Common-Cyclical Features in VAR Analysis: A Monte-Carlo Study (Preliminary Version)

Farshid Vahid, João Victor Issler

Setembro de 1999

URL: http://hdl.handle.net/10438/971

Vahid, Farshid
The Importance of Common-Cyclical Features in VAR Analysis: A Monte-Carlo Study (Preliminary Version)/ Farshid Vahid, João Victor Issler - Rio de Janeiro : FGV,EPGE, 2010
(Ensaios Econômicos; 352)

Inclui bibliografia.

CDD-330

The Importance of Common-Cyclical Features in VAR Analysis: A Monte-Carlo Study*

Farshid Vahid

Department of Econometrics and Business Statistics

Monash University

Clayton, Victoria 3168

Australia

Farshid.Vahid@BusEco.monash.edu.au

João Victor Issler

Graduate School of Economics – EPGE

Getulio Vargas Foundation

Praia de Botafogo 190 s. 1125-8

Rio de Janeiro, RJ 22253-900

Brazil

jissler@fgv.br

August 1999, preliminary version, comments are welcome.

*This paper was written while João Victor Issler was visiting Monash University, which hospitalty is gratefully acknowledged. We have bene.ted from comments and suggestions of Heather Anderson, Carlos

Abstract

Despite the belief, supported by recent applied research, that aggregate data display short-run comovement, there has been little discussion about the econometric consequences of these data "features." We use exhaustive Monte-Carlo simulations to investigate the importance of restrictions implied by common-cyclical features for estimates and forecasts based on vector autoregressive and error correction models. First, we show that the "best" empirical model developed without common cycles restrictions need not nest the "best" model developed with those restrictions, due to the use of information criteria for choosing the lag order of the two alternative models. Second, we show that the costs of ignoring common-cyclical features in VAR analysis may be high in terms of forecasting accuracy and efficiency of estimates of variance decomposition coefficients. Although these costs are more pronounced when the lag order of VAR models are known, they are also non-trivial when it is selected using the conventional tools available to applied researchers. Third, we find that if the data have common-cyclical features and the researcher wants to use an information criterium to select the lag length, the Hannan-Quinn criterium is the most appropriate, since the Akaike and the Schwarz criteria have a tendency to over- and under-predict the lag length respectively in our simulations.

1. Introduction

Common-cyclical movements in detrended economic variables have been so prevalent that they have acquired the status of "stylized facts." Lucas(1977) states that the main regularities observed in cyclical fluctuations of economic time series are in their comovement, which he itemizes as follows:

(i) Output movements across broadly defined sectors move together. (In Mitchell's

Martins-Filho, Aman Ullah, and participants of the Latin American and European Meetings of the Econometric Society of 1999, who are not responsible for any remaining errors in this paper. João Victor Issler acknowledges the support of CNPq-Brazil and PRONEX.

terminology, they exhibit high conformity; in modern time series language, they have high coherence.) (ii) Production of producer and consumer durables exhibit much greater amplitude than does the production of nondurables. (iii) Production and prices of agricultural goods and natural resources have lower than average conformity. (iv) Business profits show high conformity and much higher amplitude than other series. (v) Prices generally are procyclical. (vi) Short-term interest rates are procyclical; long-term rates slightly so. (vii) Monetary aggregates and velocity measures are procyclical.

From an empirical standpoint common cycles have been shown to be a "feature" of a variety of macroeconomic data sets. For example, 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 using GNP data for OECD countries. For US data, Issler and Vahid(1998) find common cycles for macroeconomic aggregates and Engle and Issler(1995) and Carlino and Sill(1998) find common cycles for sectoral and regional GNPs respectively. Similar to most applied macroeconomic research done in the last fifteen years, these studies investigated common-cyclical features using vector-autoregressive (VAR) or vector error-correction (VEC) models.

Although VAR and VEC models have become the "working horse" of macroeconometric studies, one of their shortcomings is the excessive number of parameters relative to the average sample size one is usually forced to work with. For example, when dealing with post-war quarterly data, and a VAR with three variables and eight lags, there are seventy five mean parameters to be estimated from about two hundred data points on each variable. Cointegration places some restrictions on VAR coefficients, especially when cointegrating vectors are brought in from economic theory; see Engle and Granger(1987). In these cases, the reduction in the number of free coefficients is not overwhelming. If the three-variable system has one known cointegrating vector, the number of free parameters reduces from seventy five to sixty nine by estimating a VEC model. Common-cyclical features can further

reduce the number of conditional-mean parameters estimated in VEC models by considering the restrictions implied by them; see Vahid and Engle(1993). If the three variables in the VEC model share one common cycle, then the number of mean parameters reduces from sixty nine to twenty seven.

The objective of this paper is to investigate the importance of restrictions implied by common-cyclical features for forecasts, impulse-response functions, and variance-decomposition of forecast errors of economic time-series based on VAR and VEC models. As far as we know, no work has studied the effects of these restrictions. However, considerable effort has been put in examining the importance of long-run comovement constraints in VAR models, especially for forecasting; see, among others, Engle and Yoo(1987), Clements and Hendry(1995), and Lin and Tsay(1996).

As shown by Engle and Yoo, the forecasting gains of imposing long-run constraints happens as the horizon gets large. In fact, in their simulations, the unconstrained VAR forecasts better than the VEC model for short horizons. Because for long horizons forecasting uncertainty gets helplessly large time-series models are most useful for forecasting in short horizons. Hence, the payoff of investigating these short-run constraints are big relative to those of investigating long-run constraints, since they may be a way of improving the effectiveness of time-series models for horizons where they are most useful.

We assess the effects of common-cyclical features on VAR and VEC models using Monte-Carlo simulations. The focus here is on the small-sample properties of the estimates of impulse-response functions and variance-decomposition of forecast errors, as well as on out-of-sample forecasting accuracy measures. We design the simulations in such a way that the results would be relevant for applied macroeconomists dealing with a limited number of data points and trying to estimate a relatively large number of parameters. To that end, we consider a variety of Data Generating Processes (DGPs) and sample sizes, which are kept close to the "typical" data applied researchers often encounter in practice.

VAR and VEC models with common cycles fall into the general category of reduced-rank multivariate models, because common serial correlation implies rank restrictions on

their parameter matrices[1]. Researchers may be reluctant to incorporate these parameter restrictions for a purely statistical reason related to the asymmetry of the consequences of over- versus under-parametrization of econometric models. One might think that failing to incorporate common-cycle restrictions when they are true will only cause inefficiency, while imposing them when they are false will cause inconsistency. Hence, it may seem wiser to live with a possibly inefficient unconstrained model rather than a misspecified inconsistent model. The fact that all models are most probably misspecified, and the large body of empirical evidence on the superior forecast performance of parsimonious models notwithstanding, this reasoning would be correct only if the empirical model that does not have common cycles built into it nests the empirical model with common cycles. However, we show that, using the average tools of an applied researcher, and the same data set, the "best" empirical model developed without common-cycle restrictions need not nest the "best" model developed with those restrictions.

The underlying reason for this rather paradoxical result has to do with the selection of lag order for the two alternative models. The common practice in VAR analysis is to use a model-selection criterium to choose the lag length. Standard model-selection criteria may find too small a lag length of reduced-rank VARs simply because this is the only possible way available to achieve parsimony. However, if the lag length and the VAR rank are chosen simultaneously, as suggested by Lütkepohl (1993, page 202), the lag length selected for reduced-rank VARs can be potentially bigger than that selected by the standard criterium. For example, the Schwarz criterium might choose a VAR(1) as the best unconstrained VAR, while the same criterium might choose a VAR(4) with one common cycle for the same data set. Obviously, a VAR(1) cannot nest a VAR(4) with common cycles. Hence, the consideration of comovement in the model selection stage may drastically alter the final model chosen.

Our simulations reveal that, when the true DGP is a reduced-rank VAR or VEC model, the lag length chosen by the standard model-selection criteria and that chosen when rank

[1] Classic references on reduced-rank VAR's include Velu, Reinsel and Wickern (1986), Ahn and Reinsel (1988), and Tiao and Tsay (1989).

and order are selected simultaneously can be quite different. Standard information criteria which place a strong penalty on overparameterization, such as the Schwarz or Hannan-Quinn criteria, may choose too small a lag-length when the true model has common cycles. However, they can improve considerably if the rank order is selected simultaneously with the lag length. We find strong evidence in favor of the ability of the Hannan-Quinn criterium in choosing the correct lag and rank order overall. Regarding the Akaike information criterium, we observe that its tendency to choose an overparameterized model when the lag order and rank are selected simultaneously worsens compared to the case when only the lag length is selected.

For horizons up to sixteen periods ahead, using several measures of forecasting accuracy, we find that forecasts produced by the "best" reduced-rank model are generally superior to those produced by the "best" model when only lag-order is selected. The same conclusions are obtained when comparing the variance decompositions of the "best" reduced rank model with that of the "best" full rank model when the sample size is 200[2]. Indeed, on average, if the Hannan-Quinn criterium is used to select lag order and rank, forecasting accuracy can be improved by up to 20%, and mean-squared-error of predicting the true variance contribution can be cut up to half in short horizons. This is a sizable effect which illustrates the potential gains associated to considering common-cyclical features whenever they exist.

The outline of the paper is as follows. Section 2 states the reduced-rank restrictions that common-cyclical fluctuations impose on parameters of VAR and VEC models, and discusses the relative merits of determining the rank order by statistical tests versus information criteria. Section 3 explains the design of the Monte-Carlo design used throughout the paper. Section 4 presents the simulation results for a small system of three variables and section 5 presents the same results for a larger system of six variables. Finally, section 6 summarizes the main conclusions of the paper.

[2] Notice that in the textbook example Lütkepohl (1993, pp. 202-3) lag selection was identical whether or not the rank was also chosen, and in that case, he observed that the forecasts and variance decompositions were quite similar for the reduced rank and full rank models.

2. Common cycles in VAR and VEC models

To match the stylized facts of most macroeconomic variables, we assume that the objective of research is to build a time series model for the growth rate of a vector of n economic variables. We denote the level of these variables at time t by Y_t, their logarithms by y_t, and their growth rates (i.e. the first difference of the logarithm of Y_t) by Δy_t. We make the reasonable assumption that Δy_t is stationary, add the assumption that Δy_t has mean zero to simplify notation, without loss of generality, and start with the Wold representation of Δy_t:

$$\Delta y_t = C(L)\varepsilon_t, \tag{2.1}$$

where $C(L) = \sum_{j=0}^{\infty} C_j L^j$ is a matrix polynomial in the lag operator L, $L^k z_t = z_{t-k}$, with $C_0 = I_n$. From the work of Beveridge and Nelson(1981) and Stock and Watson(1988) it is possible to decompose the log-level series y_t into common trends and cycles (which we refer to as the Beveridge-Nelson-Stock-Watson – or BNSW for short – decomposition). Using the identity $C(L) = C(1) + \Delta C^*(L)$, disconsidering the initial values in y_0, and integrating both sides of (2.1) we get[3]:

$$\begin{aligned} y_t &= C(1)\sum_{j=1}^{t}\varepsilon_j + C^*(L)\varepsilon_t \\ &= T_t + C_t, \end{aligned} \tag{2.2}$$

where $T_t = C(1)\sum_{j=1}^{t}\varepsilon_j$ and $C_t = C^*(L)\varepsilon_t$ stack respectively the trend and cyclical components of y_t. In the BNSW decomposition the n variables in y_t are decomposed into n random-walk components (stochastic trends) and n stationary components (stochastic cycles). If $C(1)$ has rank $n - q$ ($q > 0$), the stochastic trends in y_t can be characterized as linear combinations of only $n - q$ common random walks, in which case y_t is said to be cointegrated or have common stochastic trends, with q linearly independent cointegrating vectors stacked in the matrix β'; see Engle and Granger(1987). If $C^*(L)$ has rank r ($n - r > 0$), then the stochastic cycles in y_t can be characterized as linear combinations of r common stochastic

[3] See Stock and Watson (1988) or Vahid and Engle (1993) for more details.

cycles, with $n - r$ cofeature vectors stacked in the matrix $\beta'_\perp$; see Vahid and Engle(1993). In this paper, we investigate the costs of ignoring this singularity in the stochastic cycles C_t.

If there are $n - q$ common trends in the system, then a vector error-correction (VEC) model,

$$\begin{aligned}\Delta y_t &= A_1 \Delta y_{t-1} + \cdots + A_p \Delta y_{t-p} + \alpha \beta' y_{t-1} + \varepsilon_t \\ &= \begin{bmatrix} A_1 & \cdots & A_p & \alpha \beta' \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p} \\ y_{t-1} \end{bmatrix} + \varepsilon_t, \end{aligned} \tag{2.3}$$

would be a parsimonious representation for y_t relative to the VAR in log-levels, where the columns of the $n \times q$ matrix β are the cointegrating vectors and α is the adjustment-coefficient matrix. If there is no cointegration, then the term $\alpha \beta' y_{t-1}$ on the right-hand side of (2.3) vanishes and the model reduces to a VAR in first differences. Hence, in what follows, without loss of generality, we focus our attention on the VEC model.

If there are r common stochastic cycles in y_t, the matrix $\begin{bmatrix} A_1 & \cdots & A_p & \alpha \beta' \end{bmatrix}$, which includes all the parameters in the conditional mean of Δy_t, must have rank r. This is a consequence of the fact that $C^*(L)$ in (2.2) has rank r, which implies that there are exactly $n - r$ non-colinear linear combinations of y_t which are random walks and do not exhibit any cycles. Since the first differences of these $n - r$ linear combinations are white noise, they are unpredictable using the past. Hence, $\begin{bmatrix} A_1 & \cdots & A_p & \alpha \beta' \end{bmatrix}$ must have rank r, with a common null-space. This implies that the VEC model has itself the following parsimonious representation:

$$\begin{bmatrix} I_{n-r} & \beta'_\perp \\ 0 & I_r \end{bmatrix} \Delta y_t = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ A_1^* & \cdots & A_p^* & (\alpha \beta')^* \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p} \\ y_{t-1} \end{bmatrix} + v_t, \tag{2.4}$$

where the cofeature matrix $\tilde{e}^{0} = \begin{bmatrix} I_{n-r} & \tilde{e}^{*0} \end{bmatrix}$, which stacks the linearly-independent combinations of Δy_t that eliminates the common serial correlation in the system's components, is rotated in order to yield an $n - r$ identity sub-matrix in its first $n - r$ rows and columns, A_i^* and $(\tilde{\alpha}\tilde{\beta}')^*$ represent respectively the partitions of A_i and $\tilde{\alpha}\tilde{\beta}'$ corresponding to the remaining r equations in the system, and $v_t = \begin{bmatrix} I_{n-r} & \tilde{\beta}^{*0} \\ 0 & I_r \end{bmatrix} \varepsilon_t$. Since $\begin{bmatrix} I_{n-r} & \tilde{\beta}^{*0} \\ 0 & I_r \end{bmatrix}$ is invertible, it is possible to recover (2.3) from (2.4).

Common-cycle constraints imply important restrictions for the impulse-response functions, variance-decomposition of forecast errors, and multi-step ahead forecasts. The existence of r common cycles implies that there are $n - r$ non-collinear linear combinations of Δy_t which are white noise. Thus, from (2.1), all matrices C_i, $i = 1, 2, \ldots$ must have rank r. These matrices C_i, usually normalized to be consistent with orthogonal errors, form the basis of impulse-response functions and the components of forecast-error variance decompositions. For example, when they are post-multiplied by the cholesky factor of the variance-covariance matrix of ε_t, they yield the so-called orthogonalized impulse responses. Hence, it becomes clear that common cycles imply that the impulse responses of different variables to the same shocks are linearly dependent. Therefore, if the objects of interest are the impulse responses (or variance decompositions of forecast errors) of Δy_t, common-cycle restrictions can be used to achieve parsimony and their efficient estimation in this multivariate context.

A similar argument applies to forecasts of Δy_t at horizon h. These can be recursively calculated from:

$$\Delta y_{t+h}^{f} = A_1 \Delta y_{t+h-1}^{f} + \ldots + A_p \Delta y_{t+h-p}^{f} + \psi y_{t+h-1}^{f} \qquad (2.5)$$

where the superscript f stands for forecasts using information up to period t, $\psi = \tilde{\alpha}\tilde{\beta}'$, and actual variables are used instead of forecasts on the right-hand side where available. Since common cycles imply that the matrix $\begin{bmatrix} A_1 & \ldots & A_p & \psi \end{bmatrix}$ has reduced rank, equation (2.5) clearly shows that they will also imply that the forecasts of Δy_t at any horizon will be linearly dependent. Again, if forecasting is the objective of the multivariate model building exercise, common-cycle restrictions can also be used to achieve parsimony and their efficient

estimation.

In a VAR context above, there are two ways in which parsimony can be achieved. The first is by imposing long-run constraints (cointegration), i.e., equation (2.3), and the second is by imposing short-run constraints (common cycles), i.e., equation (2.4). The literature on forecasting has focused on long-run constraints; see, for example, Engle and Yoo(1987), Clements and Hendry(1995), and Lin and Tsay(1996). As argued in the Introduction, the payoff of investigating short-run constraints are large relative to those of investigating long-run constraints, which motivates our present research effort.

2.1. Information criteria for reduced-rank models

Our motivation is to build VAR-based models for Δy_t which can be used for forecasting, impulse-response or variance-decomposition analysis. A critical step in constructing these models is how to select the lag length of the VAR (VEC). Model-selection criteria are often used in practice to select lag length. In principle, they are useful because they do not favor any specific model against others (null versus alternative in hypothesis testing, for example); see the discussion in Granger, King and White(1995). Such criteria may choose different lag orders depending on whether or not we allow for reduced-rank parameter matrices in the VEC model[4]. Given the potential efficiency gains of using common-cycle restrictions, we investigate empirically the performance of widely-used selection criteria when only lag length is selected and when lag length and rank order are selected. Final results can then be compared to help building a strategy for empirical work.

Following Lütkepohl(1993), we focus on the Akaike, Hannan-Quinn and Schwarz information criteria for the simultaneous selection of lag and rank orders in VARs (VEC models). For the VEC model in equation (2.3), we assume that cointegrating vectors are either known

[4] Vahid and Engle(1993) focused on testing theories that implied common cycles within a VEC-model framework. They derived a statistical test for the hypothesis of r common cycles, and recommended a sequential testing procedure which could determine r. Their procedure required that the number of lags, i.e. p, be chosen in advance.

from theory, or are correctly estimated in advance. This is done here for simplicity, avoiding the well known problem of dealing simultaneously with integrated and stationary regressors in VEC models; see Toda and Phillips(1993), for example. Under this assumption, the VEC model in (2.3) can be written as:

$$\Delta y_t = \begin{bmatrix} A_1 & \cdots & A_p \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p} \end{bmatrix} + \varepsilon_t. \qquad (2.6)$$

The lag order p and the number of common cycles r, which is the rank of $\begin{bmatrix} A_1 & \cdots & A_p \end{bmatrix}$, can be simultaneously chosen to minimize one of the following model selection criteria[5]:

$$AIC(p,r) = \ln|\hat{\Sigma}_\varepsilon(p,r)| + \frac{2}{T} \cdot r \cdot (np + n - r + q) \qquad (2.7)$$

$$HQ(p,r) = \ln|\hat{\Sigma}_\varepsilon(p,r)| + \frac{2 \ln \ln T}{T} \cdot r \cdot (np + n - r + q) \qquad (2.8)$$

$$SC(p,r) = \ln|\hat{\Sigma}_\varepsilon(p,r)| + \frac{\ln T}{T} \cdot r \cdot (np + n - r + q) \qquad (2.9)$$

where q is the number of cointegrating vectors, n is the dimension of the system, r is the rank of VEC model, p is the number of lagged differences in the model, $\hat{\Sigma}_\varepsilon(p,r)$ is the estimated variance-covariance matrix of the errors of the VEC model with p lags and rank r; and T is the number of observations. Vahid and Engle(1993) showed that $r \leq q$, which implies that, given q, models of rank smaller than q need not be considered.

Calculating the information criteria in (2.7)-(2.9) for full-rank models ($r = n$) is straightforward, since they can be estimated, equation by equation, using OLS. On the other hand, for reduced-rank models, computing these criteria for different p and r may seem difficult, since one may think that their estimation is required. However, the following Lemma states a well known result that relates $\ln|\hat{\Sigma}_\varepsilon(p,r)|$ to the squared canonical correlations between

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

Δy_t and $\Delta y_{t-1} \; \Delta y_{t-2} \; \cdots \; \Delta y_{t-p} \; (\Phi D y_{t-1})'$ making computation of these criteria in this case also straightforward.

Lemma 2.1. *The minimum of* $\ln \left| \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t' \right|$ *under the assumption that* $\Pi = A_1 \; \vdots \; \cdots \; \vdots \; A_p \; \Phi \; \Psi$ *has rank* r *is*

$$\ln \left| \frac{1}{T} \sum_{t=1}^{T} \Delta y_t \Delta y_t' - \sum_{i=n-r+1}^{n} \ln(1 - \lambda_i) \right|,$$

where $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ *are the sample squared canonical correlations between* Δy_t *and the set of regressors* $x_t' \equiv \Delta y_{t-1} \; \Delta y_{t-2} \; \cdots \; \Delta y_{t-p} \; (\Phi D y_{t-1})'$. *The sample squared canonical correlations are the eigenvalues of*

$$\left(\sum_{t=p+1}^{T} \Delta y_t \Delta y_t' \right)^{-1} \left(\sum_{t=p+1}^{T} \Delta y_t x_t' \right) \left(\sum_{t=p+1}^{T} x_t x_t' \right)^{-1} \left(\sum_{t=p+1}^{T} x_t \Delta y_t' \right).$$

Proof. See Tso (1981).

This lemma shows that, after dropping the common constant term $\ln \left| \frac{1}{T} \sum_{t=1}^{T} \Delta y_t \Delta y_t' \right|$ in (2.7)-(2.9), these model-selection criteria can be expressed in terms of the eigenvalues (λ_i) as:

$$AIC(p;r) = \sum_{i=n-r+1}^{n} \ln(1 - \lambda_i(p)) + \frac{2}{T} \cdot r \cdot (np + n - r + q) \tag{2.10}$$

$$HQ(p;r) = \sum_{i=n-r+1}^{n} \ln(1 - \lambda_i(p)) + \frac{2 \ln \ln T}{T} \cdot r \cdot (np + n - r + q) \tag{2.11}$$

$$SC(p;r) = \sum_{i=n-r+1}^{n} \ln(1 - \lambda_i(p)) + \frac{\ln T}{T} \cdot r \cdot (np + n - r + q). \tag{2.12}$$

Hence, for fixed p, the model-selection criteria for any rank can be easily calculated after these eigenvalues are computed. The eigenvalues can be easily calculated by any statistical program which has a canonical correlation procedure[6]. Notice that, for fixed T, n and q, the model-selection criteria in (2.10)-(2.12) depend only on the lag length p and on the rank of the parameter matrices in the VEC model r, i.e., VEC models where the number of cointegrating vectors is known.

[6] For example, SAS or STATA, or any matrix program such as GAUSS, or by slightly modifying any of the plethora of computer programs which use this lemma to calculate the Johansen cointegration test statistics (see chapter 20 of Hamilton (1994)).

3. Monte-Carlo design

If samples are "large" and the variables have common cycles, our intuition tells us that ignoring them will not be very harmful. This is based on the expectation that with "large" samples, lag-order selection is likely to be unambiguous, and parameter estimates will be precise, so that the reduced rank constraints will be (approximately) true for the estimated parameters, even when they are not imposed at the estimation stage. Hence, the estimated models with or without common-cycle restrictions will be so close, that their results for forecasting, impulse-response, and variance-decomposition analysis will be very similar.

This intuition should not, however, be carried over to the case of "small" samples. As a matter of fact, efficiency gains are potentially relevant when sample sizes are small and degrees of freedom are scarce. In this context, using information criteria to select lag order may not be unambiguous. The standard practice is to disregard the possibility of reduced-rank models in the formula of widely-used model-selection criteria, i.e., set $r = n$ in (2.10)-(2.12). This creates the potential problem of model misspecification in selecting lag length. Indeed, selecting lag order imposing that the model is full-rank can yield a completely different result than selecting lag order and rank simultaneously. We investigate this issue using 1000 simulations of 100 reduced-rank VARs with either 100 or 200 observations each, tabulating results when lag length alone is chosen and when rank and lag length are chosen simultaneously.

To make presentation manageable, we chose to work with three- and six-dimensional VARs. In the applied macroeconomics literature, models that only consider the real side of the economy are usually three-dimensional. For example, in testing the real-business-cycle model in King, Plosser and Rebelo (1998), King et al. (1991) estimate a VAR including output, consumption, and investment. Issler and Ferreira (1998) use a VAR including output, labor, and capital inputs, to estimate long-run elasticities of the aggregate production function. Six-dimensional VARs usually include a tri-variate real-variable sub-system, as well as a monetary sub-system, including the real money supply, real interest rate, and the level of inflation; see King et al. for example.

The first parameter we set in the Monte-Carlo design is the lag length p. It is chosen in order to allow either the possibility of under- or over-parameterization of the VAR (VEC) model. This does not happen in either Lütkepohl(1985) or in some simulations in Nickels-burg(1982). The first uses a DGP with a true lag order of 1 in his simulations, making under-parameterization virtually impossible. This favors information criteria which heavily penalize over-parameterization, e.g., the Schwarz criterium[7]. The second sets the true lag order to four, but the maximum lag of four as well in the estimation stage. This makes over-parameterization impossible, favoring liberal criteria such as the AIC. To avoid both problems, for the three-dimensional system, we chose the true lag order of 4 allowing for models of up to lag 8. For the six-dimensional system, in order to save degrees of freedom, we use as the DGP a VAR with two lags and allow estimation up to lag six.

Next, we discuss the choice of variance-covariance matrix for the VAR error ε_t in the Monte-Carlo design. The properties of estimated VARs are only invariant to scaling the variance-covariance matrix of the errors by a scalar. However, the following lemma shows that in order to cover the entire space of reduced-rank VAR processes of order (p), one can fix the variance-covariance matrix of ε_t to be the identity matrix without any loss of generality.

Lemma 3.1. *Any arbitrary full rank linear transformation of a reduced-rank VAR, generates another VAR with the same rank.*

Proof. Consider a VAR(p),

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

Assume that P is a full rank $n \times n$ matrix which orthogonalizes the variance-covariance matrix of ε_t. Define $z_t \equiv Py_t$, $B_1 \equiv PA_1P^{-1}$, $\cdots$, $B_p \equiv PA_pP^{-1}$, and $\eta_t \equiv P\varepsilon_t$, where $E(\eta_t\eta_t') = I_n$. We have,

$$z_t \equiv Py_t = PA_1y_{t-1} + \cdots + PA_py_{t-p} + P\varepsilon_t$$

[7] Not surprisingly, this is exactly the criterium that does best in choosing the correct lag order in his simulations.

$$= PA_1P^{-1}Py_{t-1} + \cdots + PA_pP^{-1}Py_{t-p} + P\varepsilon_t$$
$$= B_1z_{t-1} + \cdots + B_pz_{t-p} + \varepsilon_t. \quad (3.1)$$

Since P is of full rank, the B_i's have the same rank as of their A_i's counterpart. Since all the A_i's have the same left null-space, so will all the PA_i's, and therefore so will the B_i's. ■

We now turn to choice of the coefficients in the conditional mean of the VAR (VEC) model. An exhaustive Monte-Carlo study over the entire model space is unfeasible. It is customary, as in Lütkepohl(1985), to choose several sets of eigenvalues for the companion matrix[8] of the VAR, and to choose arbitrary parameter matrices which give rise to those eigenvalues, averaging the results over all these DGPs. Although the results generated from such a design strategy might be useful for general time-series analysts, they are unsuitable for economists who work with aggregate macroeconomics data. This is because the cyclical structure (i.e. signal to noise ratio) of models including macroeconomic aggregates can be quite weak, especially for systems which do not contain a monetary sector. For example, the system R^2 (a measure similar to R^2 for univariate models which is discussed in the Appendix) for King et al.'s(1991) VEC model of US per-capita income, consumption, and investment is 0.44, whereas the system R^2 for 160 out of the 200 DGPs which Lütkepohl(1985) averages over are above 0.5, and 96 of these are greater than 0.8. Since this paper is intended primarily for applied macroeconomists, a design which gives too much weight to models with high system R^2 would be inappropriate.

Here, we start with a "typical" macroeconometric study in order to select the DGP and the system R^2 associated with it. The data set used for choosing our parameter values is the same as in King et al.(1991)[9]. For the three-variable system, we considered 100 different DGPs. Their set of parameter values are drawn randomly from the estimated asymptotic 95% confidence region of the parameters of reduced-rank VARs of order four. The latter

[8] The companion matrix of a VAR(p) is the coefficient matrix of its VAR(1) representation. The condition for VAR(p) to be stationary is that all of the eigenvalues of its companion matrix are inside the unit circle.

[9] King et al.(1991) chose a lag length of eight for their three variable model and a lag length of six for their six variable model. They chose these lag lengths on a-priori grounds, and without any reference to data.

are based on estimates of fourth-order VARs of the first-differences of the logarithms of US per-capita private income, consumption, and investment over the period 1947.1 to 1988.4, in quarterly frequency. For the six-variable model, we also considered 100 DGPs, drawned using the same method, after fitting reduced-rank VARs of order two to the first differences of logarithms of per-capita private income, consumption, investment, real money balances, interest rate and inflation over the period 1947.1 to 1988.4. For all cases, we have been careful to verify that all of these randomly drawn DGPs satisfy the stationarity conditions for Vector Autoregressions[10].

By choosing our DGPs from this "plausible" subset of the parameter space, we believe that our results would be directly relevant for applied macroeconomists. For comparison with Lütkepohl(1985), the median of the system R^2 measure for our generated three-variable DGPs is between 0.5 and 0.6, with less than 5% larger than 0.7 and none greater than 0.8.

The Monte-Carlo simulation consists of the following:

1. Using each of these 100 DGPs, we generate 1000 samples (of either 100 or 200 observations), record the lag length chosen by traditional (full-rank) AIC, HQ, and SC measures, i.e., the information criteria stated in (2.10)-(2.12) when r = n, and the lag length and rank order chosen by the information criteria stated in (2.10)-(2.12).

    1. In all cases, to reduce the impact of initial values on simulated series, we generated 500 observations. Only the last 116 or 216 observations were selected for the forecasting exercise and only the last 100 and 200 were selected for the impulse-response and variance-decomposition exercises.

2. We then compare the ability of each of these strategies of model selection in estimating the DGP's true lag length and rank. This comparison allows measuring the chance of misspecification arising from ignoring common-cycles at the lag-length selection stage.

[10] The range of the absolute value of the maximum eigenvalue of the tri-variate rank-one DGPs is $(0.49; 0.87)$. For the tri-variate rank-two DGPs this range is $(0.63; 0.92)$, and for the six-variate rank-four DGPs it is $(0.56; 0.96)$.

3. Based on the information-criteria results, a "best" model for each criterium is chosen when traditional (full-rank) criteria are used and a "best" model for each criterium is also chosen when the criteria in (2.10)-(2.12) are used. For each type of information criteria these two "best" models are compared regarding:

    1. out-of-sample forecasting accuracy up to 16 periods ahead, and,
    2. mean-squared-error in estimating variance decompositions of forecast errors and impulse-response functions.

We turn now to the specifics of the exercises on forecasting evaluation and on variance-decomposition and impulse-response function estimates.

3.1. Measuring forecasting accuracy

Appropriate evaluation of forecasts depends on the specific use that the forecasts are needed for, i.e. the "loss function" of the user. The fact that we have applied economists as our target audience does not point to any specific way that we should evaluate the forecasts of alternative models. First, it is not reasonable to impose any type of asymmetry in evaluating forecast errors. Second, a macroeconomist who models the growth rate of income, consumption and investment, might in fact be interested in the growth rates of income, savings and investment, or she might be interested in forecasting the levels based on the growth rates. Therefore, it is important to evaluate the forecasting performance of different models on the basis of measures that are invariant to linear transformation of forecasts at one horizon, or across different horizons. One measure that satisfies this invariance property is the generalized forecast error second moment (GFESM) introduced by Clements and Hendry(1994). It is the determinant of the expected value of the outer product of the vector of stacked forecast errors of all future times up to the horizon of interest. For example, if forecasts up to h

quarters ahead are of interest, this measure will be:

$$GFESM = E\left[\begin{pmatrix} \tilde{\varepsilon}_{t+1} \\ \tilde{\varepsilon}_{t+2} \\ \vdots \\ \tilde{\varepsilon}_{t+h} \end{pmatrix}' \begin{pmatrix} \tilde{\varepsilon}_{t+1} \\ \tilde{\varepsilon}_{t+2} \\ \vdots \\ \tilde{\varepsilon}_{t+h} \end{pmatrix}\right]$$

where $\tilde{\varepsilon}_{t+h}$ is the n-dimensional forecast error at horizon h of our n-variable model. It is obvious that this measure is invariant to elementary operations that involve different variables, and also to elementary operations that involve the same variable in different horizons. In the Monte-Carlo, the above expectation is evaluated for every model, by averaging over the simulations.

We also considered here two additional measures of forecasting accuracy. The first is the determinant of the mean squared forecast error matrix at different horizons ($|MSFE_h|$), and the second is the trace of the mean squared forecast error matrix ($TMSFE$). The determinant of the $MSFE$ is invariant to elementary operations on forecasts of different variables at a single horizon, but it is not invariant to elementary operations on forecasts across different horizons. The trace of the mean squared forecast error matrix is not invariant to either of these transformations.

There is one complication arising from the fact that we are simulating 100 different DGPs. In this case, the simple averaging of these measures across different DGPs is not appropriate, since 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 of many horizons. Instead, we calculate the percentage change in these measures in the "best" full-rank model and the "best" reduced-rank model chosen by each criterium for every DGP, and then average these changes over all DGPs.

3.2. Precision of impulse-response and variance-decomposition estimates

Although in many cases the objects of interest in applied studies that use VAR or VEC models are impulse-response functions and variance-decompositions of forecast errors, all simulation studies we aware of focus on forecast comparisons alone. Thus, studying the precision of estimates of impulse-response functions and variance-decomposition coefficients for different VAR and VEC models has also a high payoff.

Impulse-response functions and variance-decomposition of forecast errors differ from multi-step forecasts of VAR of VEC models in which they are not only non-linear functions of the estimates of mean-parameter matrices but of the variance-covariance matrix of system errors as well. Given this added dimension to the problem, one cannot a priori expect to get similar results to the forecasting exercise.

Moreover, there a few issues that are specific to the analysis of impulse-response functions and variance decomposition of forecast errors. First, errors have to be orthogonal for results to be meaningful[11]. As is well known, there are several techniques that yield orthogonal errors. Here, we chose to use the Choleski decomposition for the variance-covariance matrix to orthogonalize shocks, since this is by far the most popular method used in practice. Second, for a three-variable system there are nine impulse-response and variance-component coefficients in each horizon. For a six-variable system there are thirty six. In order to report results in a compact way, the mean-squared errors of each of them is computed for the rank-restricted and the unrestricted VAR model. Then, the percentage improvement in MSE of the restricted model is computed for each of these coefficients. Finally, for each horizon, the mean percentage improvement across all coefficients is computed. It should be noticed that this method ensures that final results do not depend on the unit of measurement of the variables in the system. Third, in order to keep the size of our tables down to a minimum, only variance-decomposition results are reported, since results for impulse responses followed

[11] In this case, the results in our Lemma 2 are not applicable, since Monte-Carlo results are not independent of the way chosen to orthogonalize shocks to the VAR.

a very similar pattern to those found in the variance-decomposition exercise.

3.3. A benchmark for future reference

As in any simulation study is useful to generate a benchmark case to be used for future reference. Here, we chose as a natural benchmark simulation the case where the researcher knows the true lag length of the VAR, thereby ruling out any chance of model misspeci…cation. Any di¤erences between reduced-rank and full-rank VAR models re‡ect solely e¢ ciency gains of the former.

To save space we present the results of this exercise only for the three-variable system. Table 1 shows the percentage improvement of di¤erent measures of forecast accuracy and of mean-squared error (MSE) of variance-decomposition coe¢ cients when reduced-rank VEC models are allowed for. If GFESM is considered, forecasting accuracy can improve as much as 73%, with a median improvement across all horizons and sample sizes of about 30%. If $|MSFE_h|$ and TMSFE are considered the gains are smaller (22% and 2% respectively). It also becomes clear that they happen mostly at short horizons, which does not happen when GFESM is considered, since the latter accumulates forecasting-accuracy gains across horizons. The gains in MSE of variance-decomposition coe¢ cients are also sizable - they can be as high as 66% with a median gain of about 45%, although at the …rst horizon there is a loss in MSE that can reach up to 43%.

These results show the potential e¢ ciency gains when the lag-length of VAR and VEC models are known. Although they serve as a benchmark, these gains are unrealistic for empirical studies, since there lag lengths must be estimated beforehand. We next consider reduced-rank models lag-length selection using information criteria under two distinct situations. First, when the information criteria in (2.10)-(2.12) are used setting $r = n$, and second, when they are used allowing for the possibility of reduced rank in VAR and VEC models.

This type of exercise allows investigating the following: (i) how often model selection

criteria choose a different lag length when, as is the norm in practice, only full rank models are considered? These results are then confronted with those obtained when the same criterium is used to pick the lag order and rank at the same time; (ii) do differences in the chosen models by these two classes of model selection criteria lead to major differences in forecasting accuracy? And (iii) do differences in the chosen models by these two classes of model selection criteria lead to major differences of accuracy of estimates of impulse-response and variance-decomposition coefficients. A secondary result, which can be of interest to practitioners, is the relative performance of different model-selection criteria for reduced-rank VARs in choosing their correct lag and rank order.

4. Monte-Carlo simulation results for the three-variable model

4.1. Selection of lag and rank order

Table 2.a shows the frequency of lag-order selection in 1000 simulations of 100 trivariate VARs(4) with rank 1 by AIC, HQ and SC when only full rank models are considered, and when rank and lag orders are determined simultaneously. The top half of this table corresponds to a sample size of 100, and the bottom half corresponds to samples of 200 observations. Table 2.b shows the analogous frequencies when the true DGP is a trivariate VAR(4) of rank 2.

These Tables confirm that selecting the lag and rank order jointly, can lead to choice of a model which is of higher lag-order than the lag-length chosen when only full rank models are considered. For example, the top half of Table 2.a shows that for samples of 100 observations, the modal choice of all three criteria is VAR(1), with AIC choosing the true lag of 4 only 14 percent of the time. The other two criteria choose a VAR(4) with a frequency of less than 1 percent. However, when the lag and rank are chosen simultaneously, there is a large reduction in the number of times that the VAR(1) is chosen, regardless of the criterium used. Furthermore, the frequency of the correct lag chosen increases significantly. In both Tables 2.a and 2.b, AIC chooses the correct lag and rank more often than the other two criteria,

with HQ a close second. The modal choice of the Schwarz criterium stays at a VAR(1) even with 200 observations.

Two points are worth noting. First, even in those cases in which the criteria choose the wrong lag-length, they are likely to choose the correct rank. The only exception is SC when the true rank is 2 and there are only 100 observations. This suggests that common cycles can be detected even if the wrong lag-length is chosen. This is plausible, because the property that a linear combination of variables has no dependence with the past (the necessary and sufficient condition for common cycles), is unrelated to what those cycles are and whether they are correctly specified. The second point is that AIC has a tendency to over-predict the true lag length, even when sample size is 200, once one chooses lag-length and rank simultaneously. Given the evidence on the adverse effects of overparameterization on forecasting in time series models in the literature (see Lütkepohl, 1985), this cautions us that the costs of using AIC to choose lag and rank order jointly may outweigh its benefits. The analysis of the forecasts in the next subsection confirms that this is indeed the case.

4.2. Forecasts

Tables 3.a and 3.b show the percentage improvement in these measures when lag and rank are chosen simultaneously, over when lag-length is chosen alone imposing $r = n$. A general conclusion is that reduced-rank models have no forecast ability beyond 8 periods, and most of the advantage of looking for common cycles is in forecasting one to four periods ahead. Despite that, there are still non-trivial gains for considering the possibility of reduced-rank models: GFESM can be reduced up to 32%, $|$MSFE$_h|$ up to 11% and TMSFE up to 8%, when the true rank is one. For rank two these potential reductions are respectively 31%, 9%, and 5%. These numbers are about half as large as our benchmark case when GFESM and $|$MSFE$_h|$ are considered, but are larger when TMSFE is considered.

The results in Tables 3.a and 3.b allow also comparing the three information criteria in terms of their relative performance. Regardless of the forecasting-accuracy measure used, HQ

provides the best forecasting performance across information criteria, while AIC provides the worst. Our results for HQ and SC show that there are sizable benefits from choosing lag and rank jointly. In these cases, these criteria give rise to models that are closer to the true DGP, without increasing the chance of overshooting the correct lag.

On the other hand, considering the forecast performance of AIC, and the increased possibility of over-prediction of the lag-order when lag and rank are chosen simultaneously, especially when sample size is 100, we conclude that the joint determination of lag and order by AIC in models with small R^2 is not appropriate. Indeed, if one wants to use AIC, and also wants to allow for possibility of common cycles, it would be better to employ the following strategy. First, use AIC to choose the lag length testing for common cycles using the test in Vahid and Engle(1993). Then, impose common cycles if that is not rejected by the testing procedure. In this way, the possibility of overshooting the correct lag length is somewhat controlled.

4.3. Variance-decomposition results

The percent improvement of estimates of forecast-error variance decomposition coefficients are presented in Table 4. First notice that the smallest gains (largest losses) are obtained for horizon one, where results depend exclusively on the estimate of the variance-covariance matrix of the errors[12]; a similar result is obtained in our benchmark case. This may be due to the fact that the variance contributions are a ratio of elements of the variance-covariance matrix. Hence, although the restricted model estimates the latter more precisely, it performs worse in estimating the variance ratios compared to the unrestricted model. Second, when sample size is 100 observations, there is no consistent pattern of results favoring reduced-rank criteria (i.e., when lag and rank are jointly selected). Third, increasing the sample size to 200 observations improves the gains of reduced-rank models regardless of the criterium considered. The highest realized gain happens when the HQ criterium is used in this case,

[12] Notice that, for horizon one, when the Choleski decomposition is employed, there are only five variance contributions that vary across simulations.

although this says nothing about the relative performance across criteria for estimation of variance-decomposition estimates.

5. Monte-Carlo simulation results for the six-variable model

The DGPs of the six-variable Monte-Carlo simulations are drawn uniformly from the confidence region of an estimated six-variable VAR(2) of rank 4, based on US macroeconomic aggregates. The six variables are the growth rates of per-capita income, consumption, investment and real money balances, and the first difference of interest rates and the inflation rate, from 1947:1 to 1988:4. This is the data set used in King, et al(1991). The addition of the monetary variables to the system, increases the system R^2 from 0.44 to 0.81. This is possibly due to the fact that monetary variables are more dependent on the past than real variables, and that there is a strong cross correlation between the growth rates of real money balances and income.

The maximum lag-length considered is 6, which is the lag length that King et al.(1991) used in their analysis. In order to reduce the computational costs, we have drawn only 20 DGPs and performed 500 simulations in each case. In contrast to the three-variable model, the DGPs in this exercise all have R^2 between 0.8 and 0.9; which implies that finite samples are more informative about the structure of the DGP than in the previous three variable case, and that information criteria must be more successful in selecting the correct lag. This is clear from the lag and lag-rank orders chosen by the different criteria reported in Table 5.

Table 5 shows the frequency of making the correct choice, using each of the three criteria. The problem with AIC over-shooting the correct lag, when lag and rank order is chosen simultaneously, is much more pronounced here, especially in samples of 100 observations. In this case, the regular AIC only overshoots the true lag-order in less than 2 percent, but the lag-rank version of the AIC overshoots the true lag in 9.4 percent of all of the 10,000 simulations. If we add the 8.3 percentage of times that AIC overpredicts the rank even though it predicts the lag-order correctly, we see that AIC leads to an overparameterized

model in more than 17 percent in all simulations.

The interesting information in Table 5, however, is the relative success of the Hannan-Quinn criterium in choosing the correct lag and rank order, without the risk of overparameterizing the model. This repeats what we observed in our three-variable simulations. Even with a sample size of 100, the Hannan-Quinn criterium chooses the correct lag 94.2 percent of the times, and leads to overparameterized models in less than 0.5 percent of all simulations. With a sample size of 200, the Hannan-Quinn criterium is almost perfect in choosing the correct lag-order (99.98 percent), and is the best among the three criteria for choosing the correct lag-rank combination 90.7 percent in all simulations. Although the percentage of the times that the Schwarz criterium chooses the true lag improves when the lag and rank order are chosen simultaneously, this criterium is still significantly biased towards under-parameterizing the model, even when the signal is strong and the sample size is 200.

The confusion caused by AIC when it is used to choose lag and rank order simultaneously, is reflected in its forecast performance. Table 6 shows the improvement in the forecasting performance of models chosen by the different criteria when rank order is chosen in the model selection stage over when it is ignored. When the sample size is 100, using AIC to choose the lag-length and rank order simultaneously leads to very poor forecast performance of the selected model, for longer than one step ahead horizons. Given the relative success of the Hannan-Quinn criterium in choosing the correct lag-rank order, it leads to models with the best overall forecast performance, with AIC being a close second only when the sample size is 200. Unlike the case where the signal to noise ratio was small, the underparameterization of the true model by the Schwarz criterium is reflected by a very poor relative forecast performance of models chosen by this criterium.

6. Conclusions

This paper argues that the stylized fact that "macroeconomic variables move together over the business cycle" should be taken seriously in econometric models using them, usually

a VAR or a VEC model. The Monte-Carlo analysis in this paper suggests the following messages to practitioners who analyze data which exhibit comovement:

1. There are non-trivial gains in forecasting accuracy and in MSE of estimates of variance-decomposition coefficients when we allow VAR and VEC models to be of reduced rank. These results are stronger in our benchmark case (lag order is known), but are confirmed as well for the more realistic case where lag order is selected using popular information criteria.

2. Information criteria that allow for reduced-rank VAR and VEC models perform better in choosing the correct lag length compared to these same information criteria that disregard reduced-rank models ($n = r$).

3. If HQ and SC are used in a context which allows them to pick reduced rank models, then the problem that they underpredict the true lag-length is significantly remedied.

4. Do not use AIC to choose lag and rank order at the same time, especially when the sample size is small.

5. The Hannan-Quinn criterium seems to be the best criterium for choosing lag-length and rank order at the same time.

The analysis also confirms that AIC can lead to over-parameterized models with adverse consequences for forecast performance. However, it reveals that Lütkepohl's(1985) conclusion that models selected by Schwarz criterium lead to best forecasts is an artifact of his Monte-Carlo design.

References

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.

Beveridge, S. and Nelson, C.R. (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.

Carlino, G. and K. Sill (1998), "Common trends and common cycles in regional per-capita incomes", Working Paper, Federal Reserve Bank of Philadelphia.

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

Clements, M. P. and D. F. Hendry (1994), "Towards a theory of economic forecasting", Chapter 2 in C. P. Hargreaves (Ed.), Nonstationary Time Series Analysis and Cointegration, Oxford University Press.

Engle, R. F. and S. J. Brown (1986), "Model selection for forecasting", Applied Mathematics and Computation, 20, 313-327.

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

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

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

Granger, C. W. J., M. L. King and H. White (1995), "Comments on testing economic theories and the use of model selection criteria", Journal of Econometrics, 67, 173-187.

Hamilton, J. D. (1994), Time Series Analysis, Princeton University Press.

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

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

Issler, J.V. and Ferreira, P.C.(1998), "Time-Series properties and empirical evidence of growth and infrastructure," Working Paper # 335, Graduate School of Economics, Getulio Vargas Foundation.

Issler, J.V. and Vahid, F.(1998), "Common cycles and the importance of transitory shocks to macroeconomic aggregates," Working Paper # 334, Graduate School of Economics, Getulio Vargas Foundation.

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

Lucas, R. E. Jr.(1977), "Understanding business cycles", Carnegie-Rochester Series on Public Policy, 5, 7-29, also reprinted in Lucas, R. E. Jr. (Ed) Models of Business Cycle, 1977.

Lütkepohl, H.(1993), Introduction to Multiple Time Series Analysis, Second Edition, Springer-Verlag.

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

Magnus, J.R. and H. Neudecker(1988), Matrix Differential Calculus with Applications in Statistics and Econometrics, John Wiley.

Neusser, Klaus(1991). "Testing the Long-Run Implications of the Neoclassical Growth Model," Journal of Monetary Economics, v27(1), 3-38.

Nickelsburg, G.(1982), "Small-sample properties of dimensionality statistics for fitting VAR models to aggregate economic data: A Monte Carlo study", Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 155-160.

Reinsel, G. C. (1993), Elements of Multivariate Time Series Analysis, Springer-Verlag.

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

Toda, Hiro Y. and Peter C. B. Phillips (1993), "Vector Autoregressions and Causality," Econometrica, vol. 61 (6), pp. 1367-1393.

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

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

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

Velu, R. P., G. C. Reinsel and D. W. Wickern (1986), "Reduced rank models for multiple time series", Biometrika, 73, 105-118.

A. System R^2 and signal-to-noise ratio

In a multiple regression with stochastic regressors and i.i.d. errors, $y = X\beta + \varepsilon$; the limiting signal-to-noise ratio (snr) can be defined as:

$$snr = \frac{\beta' \lim_{T\to\infty} E\left[\frac{X'X}{T}\right]\beta}{\sigma^2_\varepsilon}; \qquad (A.1)$$

where $E(\varepsilon\varepsilon') = \sigma^2_\varepsilon I$, and the proportion of the variation of dependent variable explained by the model, i.e. the population R^2, is:

$$R^2 = \frac{\beta' \lim_{T\to\infty} E\left[\frac{X'X}{T}\right]\beta}{\sigma^2_\varepsilon + \beta' \lim_{T\to\infty} E\left[\frac{X'X}{T}\right]\beta} = \frac{snr}{(1+snr)}.$$

Since the asymptotic variance of $\sqrt{T}(\hat{\beta} - \beta)$ is $AVAR(\hat{\beta}) = \sigma^2 \lim_{T\to\infty} E\left(\frac{X'X}{T}\right)^{-1}$, we can write (A.1) as:

$$snr = \beta' AVAR(\hat{\beta})^{-1}\beta. \tag{A.2}$$

Consider now a VAR(p):

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t. \tag{A.3}$$

The analogous measure of snr for it is:

$$snr = \gamma' \Sigma_{\Gamma\Gamma}^{-1} \gamma \tag{A.4}$$

where $\gamma = vec(A)$, $A = [A_1 \; \cdots \; A_p]$, $E[\varepsilon_t \varepsilon_{t-j}'] = \Gamma_j$, and:

$$\Sigma_{\Gamma\Gamma} = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & \cdots & \Gamma_{p-2} \\ \cdots & \cdots & \cdots & \cdots \\ \Gamma_{p-1}' & \Gamma_{p-2}' & \cdots & \Gamma_0 \end{bmatrix};$$

where $\Gamma_j = E[y_t y_{t-j}']$. Notice that $\Sigma_{\Gamma\Gamma}$ is completely determined by (A, Γ) via the Yule-Walker equations[13]. After some algebra, it can be shown that (A.4) is equal to:

$$snr = \gamma' \Sigma_{\Gamma\Gamma}^{-1} \gamma = trace[\Gamma_0 \Gamma^{-1}\Gamma].$$

Using this last result, one can then define the system R^2 to be:

$$R^2 = \frac{trace(\Gamma_0 \Gamma^{-1})}{1 + trace(\Gamma_0 \Gamma^{-1})}.$$

[13] See Hamilton (1994) Chapter 10, Lütkepohl (1993) Chapter 1, or Reinsel (1993) Chapter 2.

Table 1: Percentage improvement in MSE of forecast-error variance decompositions, and different forecast measures when the lag length is known

horizon	True rank is one				True rank is two			
(h)	GFESM	JMSFE$_h$j	TMSFE	Var. Dec.	GFESM	JMSFE$_h$j	TMSFE	Var. Dec.
Sample size 100								
1	22.22	22.22	1.97	-14.10	12.08	12.08	0.98	-39.60
4	60.41	8.66	1.72	45.55	34.56	5.02	0.96	45.23
8	70.54	1.70	1.39	46.91	42.57	1.32	0.77	43.84
12	72.34	0.46	1.03	46.25	44.66	0.45	0.59	46.91
16	72.86	0.21	0.81	45.95	45.26	0.25	0.47	47.73
Sample size 200								
1	11.22	11.22	1.14	-7.92	6.55	6.55	0.60	-42.53
4	29.53	3.80	0.88	64.50	17.80	2.32	0.52	9.32
8	32.97	0.52	0.64	66.20	20.72	0.46	0.37	13.18
12	33.37	0.11	0.45	64.07	21.19	0.07	0.27	14.51
16	33.47	0.05	0.34	63.52	21.25	0.06	0.20	14.90

Table 2.a: Frequency of lag (p) and lag-rank (p; r) choice by different criteria when the true models are (4;1)

Selected lag	1			2			3			4			5			6			7			8		
Selected rank	1	2	3	1	2	3	1	2	3	1^T	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Number of observations= 100																								
AIC (p)	i	i	57.0	i	i	13.1	i	i	12.6	i	i	14.0	i	i	2.0	i	i	0.7	i	i	0.3	i	i	0.3
AIC (p; r)	10.8	2.5	0.4	7.4	2.0	0.1	14.4	2.4	0.1	32.7	3.4	*	8.3	1.1	*	5.0	0.6	*	3.8	0.4	*	4.0	0.5	*
HQ (p)	i	i	92.9	i	i	4.7	i	i	1.7	i	i	0.7	i	i	*	i	i	*	i	i	0	i	i	0
HQ (p; r)	39.2	1.9	0.2	13.3	0.3	*	17.0	0.1	*	24.3	0.1	*	2.4	*	0	0.7	*	0	0.3	0	0	0.1	0	0
SC (p)	i	i	99.6	i	i	0.4	i	i	*	i	i	*	i	i	0	i	i	0	i	i	0	i	i	0
SC (p; r)	73.8	0.4	*	10.7	*	0	8.4	0	0	6.6	0	0	0.1	0	0	*	0	0	0	0	0	0	0	0
Number of observations= 200																								
AIC (p)	i	i	25.9	i	i	10.7	i	i	20.0	i	i	40.0	i	i	2.7	i	i	0.5	i	i	0.2	i	i	*
AIC (p; r)	2.2	0.7	0.1	3.3	0.8	*	12.1	1.8	*	56.4	4.1	0.1	9.1	0.8	*	4.1	0.3	0	2.3	0.1	0	1.6	0.1	0
HQ (p)	i	i	80.1	i	i	7.8	i	i	7.2	i	i	4.9	i	i	*	i	i	0	i	i	0	i	i	0
HQ (p; r)	16.1	0.6	0.1	8.9	0.1	*	20.7	0.1	0	51.3	*	0	1.9	0	0	0.2	0	0	*	0	0	*	0	0
SC (p)	i	i	98.6	i	i	1.0	i	i	0.3	i	i	0.1	i	i	0	i	i	0	i	i	0	i	i	0
SC (p; r)	49.4	0.1	*	11.1	*	0	17.1	0	0	22.3	0	0	0.1	0	0	*	0	0	0	0	0	0	0	0

Numbers in each cell represent percentage times that the model selection criterion corresponding to that row chose the lag-rank order corresponding to that column in 100,000 simulations (1000 simulations of 100 different DGPs). The true lag-order is identified with superscript T. A * corresponds to a non-zero value less than 0.05 percent. Numbers in a row may not add up to 100.0 exactly because of rounding.

Table 2.b: Frequency of lag (p) and lag-rank (p; r) choice by di¤erent criteria when the true models are (4;2)

Selected lag	1			2			3			4			5			6			7			8		
Selected rank	1	2	3	1	2	3	1	2	3	1	2^T	3	1	2	3	1	2	3	1	2	3	1	2	3
N umber of observations= 100																								
A IC (p)	i	i	19.9	i	i	10.2	i	i	21.3	i	i	41.3	i	i	4.6	i	i	1.5	i	i	0.7	i	i	0.5
A IC (p; r)	1.1	4.9	1.0	1.0	4.7	0.6	2.5	15.5	1.2	4.3	43.7	1.8	1.2	7.0	0.3	0.8	3.1	0.1	0.7	1.8	*	0.9	1.6	*
HQ (p)	i	i	64.1	i	i	13.1	i	i	12.7	i	i	9.9	i	i	0.1	i	i	*	i	i	0	i	i	0
HQ (p; r)	8.6	19.6	1.9	5.0	8.1	0.2	8.1	14.8	0.1	10.5	20.8	*	1.1	0.6	0	0.4	0.1	0	0.1	*	0	0.1	*	0
SC (p)	i	i	93.2	i	i	5.1	i	i	1.5	i	i	0.2	i	i	0	i	i	0	i	i	0	i	i	0
SC (p; r)	30.3	30.6	1.2	9.5	4.8	*	9.4	4.3	*	7.9	1.9	0	0.2	*	0	*	0	0	*	0	0	0	0	0
N umber of observations= 200																								
A IC (p)	i	i	3.3	i	i	2.7	i	i	16.3	i	i	72.2	i	i	4.3	i	i	0.8	i	i	0.2	i	i	0.1
A IC (p; r)	0.1	0.6	0.2	0.1	0.9	0.1	0.3	10.2	0.7	0.9	72.3	2.5	0.2	7.1	0.2	0.1	2.0	*	0.1	0.8	*	0.1	0.4	*
HQ (p)	i	i	27.9	i	i	9.6	i	i	23.3	i	i	39.2	i	i	*	i	i	0	i	i	0	i	i	0
HQ (p; r)	1.3	7.5	0.6	0.9	4.7	*	3.4	20.0	*	4.7	56.2	*	0.2	0.4	0	*	*	0	*	0	0	*	0	0
SC (p)	i	i	74.4	i	i	10.4	i	i	10.3	i	i	5.0	i	i	0	i	i	0	i	i	0	i	i	0
SC (p; r)	9.2	26.9	0.7	4.3	6.8	*	8.2	15.1	0	9.1	19.8	0	*	*	0	*	0	0	0	0	0	0	0	0

N umbers in each cell represent percentage times that the model selection criterion corresponding to that row chose the lag-rank order corresponding to that column in 100,000 simulations (1000 simulations of 100 di¤erent D G P s). T he true lag-order is identi…ed with superscript T . A * corresponds to a non-zero value less than 0.05 percent. N umbers in a row may not add up to 100.0 exactly because of rounding.

Table 3a: Percentage improvement in different measures of accuracy in forecasts generated by the best possibly reduced rank VAR over the best full rank VAR chosen by the same model selection criterion when the true models are trivariate (4,1)

horizon (h)	AIC			HQ			SC		
	GFESM	$\|MSFE_h\|$	TMSFE	GFESM	$\|MSFE_h\|$	TMSFE	GFESM	$\|MSFE_h\|$	TMSFE
Sample size 100									
1	6.6[w]	6.6[w]	0.0[w]	6.8[b]	6.8[b]	2.8[b]	5.3	5.3	1.6
4	10.8[w]	2.3	1.1	16.1[b]	6.1[b]	4.8[b]	10.9	4.1[w]	3.0[w]
8	4.0	-1.0	0.0	15.1	-0.3	2.7	11.0	-0.1	1.7
12	2.0	-0.6	-0.2	14.2	-0.2	1.7	10.7	-0.1	1.1
16	1.0	-0.3	-0.2	13.7	-0.2	1.2	10.5	-0.1	0.8
Sample size 200									
1	9.1	9.1	2.0	11.0	11.0	6.7	8.3	8.3	5.3
4	22.2	3.2	2.0	30.8	8.2	7.7	22.5	7.1	6.8
8	22.1	0.1	1.0	31.8	0.5	4.4	23.4	0.4	3.9
12	22.1	0.0	0.7	31.7	0.0	2.8	23.4	0.0	2.6
16	22.0	0.0	0.5	31.7	0.0	2.1	23.3	0.0	1.9

GFESM is Clements and Hendry's generalized forecast error second moment measure, $|MSFE_h|$ is the determinant of the mean squared forecast error matrix for horizon h, TMSFE is the trace of the MSFE matrix. Superscripts b and w denote the best and the worst performance among the three information criteria.

Table 3b: Percentage improvement in different measures of accuracy in forecasts generated by the best possibly reduced rank VAR over the best full rank VAR chosen by the same model selection criterion when the true models are trivariate (4,2)

horizon	AIC			HQ			SC		
(h)	GFESM	$\|MSFE_h\|$	TMSFE	GFESM	$\|MSFE_h\|$	TMSFE	GFESM	$\|MSFE_h\|$	TMSFE
Sample size 100									
1	7.6	7.6	0.1	5.9	5.9	2.2	1.6	1.6	1.4
4	19.2	2.9	0.5	19.2	6.1	3.9	10.1	6.1	4.3
8	20.4	0.1	0.1	19.7	-0.0	2.2	10.0	-0.0	2.5
12	20.5	0.0	0.0	19.6	-0.1	1.4	9.4	-0.0	1.6
16	20.5	0.1	0.0	19.4	-0.1	1.0	9.1	-0.1	1.2
Sample size 200									
1	5.9	5.9	0.7	6.8	6.8	2.3	8.8	8.8	5.4
4	15.3	2.0	0.5	20.5	4.3	2.6	28.7	8.9	6.5
8	17.1	0.2	0.3	21.7	0.3	1.5	31.1	0.6	3.7
12	17.3	0.0	0.2	21.8	0.0	1.0	31.3	0.1	2.5
16	17.3	0.0	0.1	21.7	0.0	1.0	31.2	-0.0	1.8

GFESM is Clements and Hendry's generalized forecast error second moment measure, $|MSFE_h|$ is the determinant of the mean squared forecast error matrix for horizon h, TMSFE is the trace of the MSFE matrix. Superscripts b and w denote the best and the worst performance among the three information criteria.

Table 4: Percentage improvement in MSE of forecast-error variance decomposition generated by the best possibly reduced rank VAR over the best full rank VAR chosen by the same model selection criterion

horizon	True rank is one			True rank is two		
(h)	AIC	HQ	SC	AIC	HQ	SC
Sample size 100						
1	-20.99	-5.51	1.47	-14.39	-8.98	-9.26
4	-5.68	1.17	7.20	13.04	5.41	-8.61
8	-8.00	-0.85	4.87	8.87	2.08	-10.77
12	-9.26	-2.00	4.20	7.97	1.42	-11.14
16	-9.79	-2.44	3.96	7.56	1.19	-11.28
Sample size 200						
1	-8.10	12.37	20.25	-4.45	0.52	9.68
4	29.15	50.65	40.91	27.15	32.84	21.20
8	27.40	57.51	38.60	24.75	31.37	20.17
12	25.79	56.30	37.41	24.06	31.03	19.73
16	25.35	55.81	37.05	23.84	30.90	19.56

Table 5: Frequency of lag (p) and lag-rank (p;r) choice by different criteria when the true model is a six variable VAR (2,4)

Selected lag	1			2			3-6		
Selected rank	rank< 4	rank= 4	rank> 4	rank< 4	rank= 4^T	rank> 4	rank< 4	rank= 4	rank> 4
Number of observations= 100									
AIC (p)	i	i	1.2	i	i	97.1	i	i	1.7
AIC (p;r)	0.1	0.1	*	9.3	72.8	8.3	2.7	6.3	0.5
HQ (p)	i	i	25.2	i	i	74.8	i	i	0
HQ (p;r)	4.3	1.4	0.1	40.6	53.2	0.4	0.1	*	0
SC (p)	i	i	77.5	i	i	22.5	i	i	0
SC (p;r)	32.0	2.8	*	52.5	12.8	0	0	0	0
Number of observations= 200									
AIC (p)	i	i	0	i	i	99.8	i	i	0.2
AIC (p;r)	0	0	0	0.6	89.5	8.5	*	1.2	0.2
HQ (p)	i	i	0.8	i	i	99.2	i	i	0
HQ (p;r)	*	0	0	9.0	90.7	0.2	0	0	0
SC (p)	i	i	28.6	i	i	71.4	i	i	0
SC (p;r)	3.2	1.2	0	40.4	55.2	0	0	0	0

Numbers in each cell represent percentage times that the model selection criterion corresponding to that row chose the lag-rank order corresponding to that column in 10,000 simulations (500 simulations of 20 different DGPs). The true lag-order is identified with superscript T. A * corresponds to a non-zero value less than 0.05 percent. Numbers in a row may not add up to 100.0 exactly because of rounding.

Table 6 Percentage improvement in different measures of accuracy in forecasts when the true model is a six variable VAR (2,4)

horizon (h)	AIC			HQ			SC		
	GFESM	$\|MSFE_h\|$	TMSFE	GFESM	$\|MSFE_h\|$	TMSFE	GFESM	$\|MSFE_h\|$	TMSFE
Sample size 100									
1	1.7	1.7	-1.2	14.1	14.1	1.1	15.4	15.4	4.4
4	-6.6	-2.2	-0.3	25.0	3.2	1.3	30.2	6.0	2.4
8	-15.4	-1.0	-0.3	25.1	0.8	0.7	24.6	2.4	1.2
12	-18.2	-0.7	-0.2	25.4	0.7	0.5	21.7	0.7	0.8
16	-18.3	-0.8	-0.2	25.5	0.1	0.4	20.8	-0.4	0.6
Sample size 200									
1	5.8	5.8	0.1	6.8	6.8	0.3	13.7	13.7	2.5
4	11.6	1.2	0.4	12.9	1.4	0.4	25.8	3.8	1.4
8	12.2	0.3	0.2	13.7	0.3	0.3	26.2	1.5	0.7
12	12.1	0.1	0.2	13.6	0.2	0.2	25.7	0.6	0.5
16	12.2	0.0	0.1	13.8	0.0	0.1	25.2	0.2	0.4

GFESM is Clements and Hendry's generalized forecast error second moment measure, $|MSFE_h|$ is the determinant of the mean squared forecast error matrix for horizon h, TMSFE is the trace of the MSFE matrix. Superscripts b and w denote the best and the worst performance among the three information criteria.