ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

VARs, Cointegration and Common Cycle Restrictions

Heather M. Anderson and Farshid Vahid

May 2010

Working Paper 14/10

VARs, Cointegration and Common Cycle Restrictions

Heather M. Anderson
Department of Econometrics and Business Statistics
Monash University
Clayton, Vic 3800
Australia

Farshid Vahid
Department of Econometrics and Business Statistics
Monash University
Clayton, Vic 3800
Australia
April 2010

Abstract:

This paper argues that VAR models with cointegration and common cycles can be usefully viewed as observable factor models. The factors are linear combinations of lagged levels and lagged differences, and as such, these observable factors have potential for forecasting. We illustrate this forecast potential in both a Monte Carlo and empirical setting, and demonstrate the difficulties in developing forecasting "rules of thumb" for forecasting in multivariate systems.

Keywords: Common factors, Cross equation restrictions, Multivariate forecasting, Reduced rank models.

JEL classification C32, C53, E37.

1 Introduction

The influential paper of Sims (1980) highlighted the importance of data-consistent multivariate dynamic specification for macroeconomic analysis, and indirectly brought the advances in dynamic econometrics and multivariate time series to the attention of macroeconomists. One significant impact of Sims' work was that it made the vector autoregressive (VAR) model the starting point of macroeconometric modelling. VAR models do not presume that any of the variables are exogenous. All equations in a VAR share the same set of regressors, making estimation simple because the full system estimator is the same as the single equation OLS estimator.

A problem with VARs is their large number of parameters. An n-variable VAR with p lags of every variable and a constant in each equation has a total of $n + n^2p$ parameters. This necessitates restricting the analysis to small systems (small n) and being frugal with the number of lags in the system (small p). Sims noted this problem and recommended the use of a Bayesian approach to control the sample variability of parameter estimates of a VAR. Since then, there have been important advances in the specification and estimation of Bayesian VARs and their application to macroeconomic forecasting (see, among others, Doan et al. (1984), Litterman (1986), Kadiyala and Karlsson (1997), Sims and Zha (1998), Waggoner and Zha (1999) and Koop et al. (2009)).

There are other ways that the problems associated with having a large number of parameters can be addressed. For example, one can allow for different lag length in each equation and determine the lag structure in each equation separately. Alternatively one can consider parameter restrictions that imply Granger non-causality or lack of feedback from one variable to another. In this chapter we follow another strand of literature which addresses the problem of over-parameterisation in VARs by searching for and imposing cross-equation restrictions that have theoretical appeal and are supported by the data.

In the absence of cross-equation restrictions, VARs are merely a collection of autoregressive distributed lag models. Even when some equations have different right hand side variables, given the uncertainty surrounding the specification of each equation, the superiority of system estimation over single equation estimation is doubtful. In such situations, the only advantage of thinking about this collection of single equations as a vector model is the trivial convenience of their vector form in multi-step forecasting. Cross-equation restrictions provide an important motivation to consider approaches that are inherently multivariate. Such approaches include the use of canonical correlation analysis (CCA) and partial canonical correlation analysis (PCCA) for uncovering general linear restrictions in a VAR. While the origins of the application of CCA and PCCA to multivariate time series modelling dates back at least to Akaike (1976) and Box and Tiao (1977), the popularity of these methods in macroeconometric modelling has risen significantly since Johansen (1988) first used them for cointegration analysis.

Johansen (1988) suggested a test for cointegration based on the canonical correlations between the vector of the first differences of I(1) variables and the vector of lagged level of these variables, after the influence of lagged differences and deterministic components have been partialled out. While the asymptotic distribution of the test statistic for the significance of canonical correlations in the I(1) case is different from that in the stationary case, the mechanics of the analysis is the same. Moreover, under the additional assumption that the true model is a finite order VAR in levels with normally distributed errors, Johansen (1988) showed that the canonical variates are maximum likelihood estimators of the cointegrating vectors. Even without such extreme assumptions, canonical variates are consistent estimators of the cointegrating vectors that are more efficient than estimators generated from static regressions.

Cointegration implies that variables in a system have common I(1) trends, and this implication is often consistent with economic theory. Two well known examples are the theory of real business cycles and the theory of the term structure of interest rates. Further, the Granger Representation Theorem (Engle and Granger (1987)) implies that cointegrated variables have a vector error-correction (VEC) representation, and the associated VECMs have become popular with economists because the error-correction terms model deviations from long run relationships between the variables.

Theoretical macroeconomic models often imply that the fluctuations in the deviations of macroeconomic variables from their equilibrium paths are governed by a small number of cyclical factors. The similarity of fluctuations of some economic variables over the course of business cycles (their pro or counter cyclicality) has acquired the status of a stylised fact. Vahid and Engle (1993) derived the implications of common cycles for a Beveridge-Nelson decomposition of integrated variables and showed that in the context of VARs these implications necessitate certain rank restrictions on the parameter matrices of the VAR. These restrictions can be tested within a canonical correlations framework, and this relates common cycles analysis to the literature on reduced rank VARs, notably Velu et al. (1986), Ahn and Reinsel (1988) and Reinsel and Ahn (1992).

In this chapter we argue that VAR models with cointegration and common cycles (or weaker forms of rank restrictions) can be usefully viewed as observable factor models. The factors are linear combinations of lagged levels and lagged differences, and as such, these observable factors have potential for forecasting. We illustrate this forecast potential in both a Monte Carlo and empirical setting, and demonstrate the difficulties in developing forecasting "rules of thumb" for forecasting in multivariate systems.

The chapter outline is as follows. Section 2 provides a synopsis of the literature on VARs with common trends, common cycles and other common features. This includes subsections on model specification, the effects of cointegration on forecasting, the use of cross equation restrictions to allow for common structural shifts and other common nonlinear features, and the possible application of shrinkage to reduced rank VECMs. The use of time varying parameter models for forecasting is also discussed at this point. Section 3 extends the Monte Carlo analysis in Lin and Tsay (1996) to illustrate how model selection and the imposition of short and long-run restrictions affect forecasts. Section 4 studies the forecasting performance of several reduced rank multivariate models of an updated version of the Litterman (1986) data set and Section 5 concludes.

2 Cointegration, common cycles and VARs

In this section we discuss cointegration and common cycles for a set of difference stationary variables before we restrict attention to VARs and forecasts derived from VARs. This is quite deliberate, because we want to emphasise that cointegration and common cycles describe common statistical properties or "common features" (Engle and Kozicki (1993), Vahid (2006)) in time series that are not restricted to VAR data generating processes. Finite order VARs are convenient approximations or filtering devices that lead to convenient asymptotic tests for such common features. Moreover, parsimonious finite order VARs are convenient forecasting tools for multivariate time series.

We consider an n-dimensional vector of non-seasonal I(1) variables. By definition, the vector of their first differences has a Wold representation

$$\Delta y_t = \mu + \Theta(L)\,\varepsilon_t,$$

where $\Theta(L)$ is an infinite moving average matrix polynomial in the lag operator with $\Theta_0 = I_K$ and absolutely summable coefficients, and ε_t are innovations in Δy_t (and equivalently innovations in y_t). Using the matrix identity $\Theta(L) = \Theta(1) + (1-L)\,\Theta^*(L)$ and integrating leads to the multivariate version of the Beveridge and Nelson (1981) decomposition of I(1) series

$$y_t = y_0 + \mu t + \Theta(1)\sum_{i=1}^{t}\varepsilon_i + \Theta^*(L)\,\varepsilon_t. \qquad (1)$$

As is customary in time series analysis, we can think of y_t as being composed of the sum of n I(1) "trends" and n I(0) "cycles",

$$\underset{n\times 1}{y_t} = \underset{n\times 1}{trend_t} + \underset{n\times 1}{cycle_t}.$$

The single distinction that trends are I(1) and cycles are I(0) does not lead to an identifiable structure, because any arbitrary set of n I(0) variables taken away from y_t leads to a set of n I(1) "trends". The additional requirement that trends are random walks (with drift when appropriate) makes the trend-cycle representation more concrete. The Beveridge-Nelson (BN) decomposition is one such decomposition in which stochastic trends and cycles are driven by ε_t, and for this reason the BN decomposition is closely affiliated with "single source of error" models (Anderson et al. (2006)). Since the BN decomposition is always possible, trend-cycle decompositions of I(1) variables always exist but may not be unique.

If $trend_t$ can be written as a linear combination of $n - q$ common I(1) factors, then y_t will be cointegrated because there will be q combinations of y_t that will be I(0). Analogously, if $cycle_t$ can be written as a linear combination of r $(< n)$ common cycles, then there will be $n-r$ linear combinations of y_t that are random walks, which implies that there are $n - r$ linear combination of $\Delta\, y_t$ that are not predictable. We emphasise that these statistical implications of common trends and common cycles are quite general, and are not specific to finite VAR processes, or BN-trends and cycles, or unobserved component models with a single source of error. However, since all observationally equivalent unobserved component structures lead to the same best linear predictor for y_t and therefore produce the same forecast, in the forecasting context, we can concentrate on the BN representation without loss of generality.

In the BN representation, $n - q$ common trends imply that there is an $n \times q$

matrix β of full column rank such that $\beta' y_t$ is I(0), i.e.

$$\beta' \mu = 0, \text{ and } \beta' \Theta(1) = 0.$$

It is also possible to separate the deterministic and the stochastic components of the trend and consider a case where $\beta' \Theta(1) = 0$ but $\beta' \mu \neq 0$, i.e., there are linear combinations of y_t that are trend stationary. The Granger representation theorem (Engle and Granger (1987)) shows that every system of cointegrated I(1) variables has a vector error-correction (VEC) representation. More specifically, in the context of a finite order VAR

$$y_t = c + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots + \pi_p y_{t-p} + \varepsilon_t, \tag{2}$$

this theorem implies that in the error-correction representation of the system

$$\Delta y_t = c + \Pi y_{t-1} + \pi_1^* \Delta y_{t-1} + \cdots + \pi_{p-1}^* \Delta y_{t-p+1} + \varepsilon_t,$$

where $\Pi = -(I - \pi_1 - \cdots - \pi_p)$ and $\pi_i^* = -\sum_{j=i+1}^{p} \pi_j$, the matrix Π has rank q and can be written as $\alpha\beta'$ where α and β are $n \times q$ matrices of rank q and β are the cointegrating vectors. Defining $w_{t-1} \doteq (\Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')'$ and $\Phi \doteq \left(\pi_1^* \vdots \cdots \vdots \pi_{p-1}^*\right)$, the vector error-correction model (VECM) can be compactly written as

$$\Delta y_t = c + \Pi y_{t-1} + \Phi w_{t-1} + \varepsilon_t. \tag{3}$$

Johansen (1988) shows that as in classic reduced rank regressions studied by Anderson (1951), the rank of Π can be determined by examining the number of statistically insignificant partial canonical correlations[1] between Δy_t and y_{t-1}, after removing the linear influence of the constant and w_{t-1} from both. In this setting, the asymptotic distributions of the likelihood ratio test statistics are no longer chi-squared. Moreover, the distribution changes depending on whether the cointegration hypothesis includes $\beta'\mu = 0$ as well as $\beta'\Theta(1) = 0$, or whether the estimated model includes a constant and other exogenous variables or not. For a thorough exposition of the likelihood ratio tests for cointegration, we refer the reader to Johansen (1995).

Similarly, r common cycles in (1) imply that there is an $n \times (n-r)$ matrix $\widetilde{\alpha}$ of full column rank such that

$$\widetilde{\alpha}'\Theta^*(L) = 0.$$

This means that there are $(n-r)$ linearly independent combinations of y_t that are combination of the trend components only, which in turn implies that the same combinations of Δy_t are linear combinations of innovations (ε_t) only, and

[1] We have provided a brief description of canonical correlations analysis in the Appendix, and refer the readers to Hamilton (1994) for more details.

hence they are unforecastable from the past information. For the VECM in (3), this implies that $\widetilde{\alpha}'\Pi = 0$ and $\widetilde{\alpha}'\Phi = 0$, i.e. the matrix $\left(\Pi \vdots \Phi\right)$ has rank r. If the rank of Π is q, the rank of this ensemble matrix is at least q. Vahid and Engle (1993) show that conditional on q and $\hat{\beta}$ – for some estimate of β that converges to β at a faster rate than $\sqrt{T}$ – a canonical correlation analysis between Δy_t and $\left(z'_{t-1}, w'_{t-1}\right)'$ with $z_{t-1} = \hat{\beta}' y_{t-1}$ can be used to determine the number of common cycles. The test statistic will have the usual chi-squared distribution. The canonical correlation computations in the Johansen procedure deliver such $\hat{\beta}$. Conditional on this, the second stage canonical correlations analysis delivers a set of weights for constructing r linear combinations of z_{t-1} and w_{t-1} that can fully explain all serial correlation in Δy_t. Moreover, these are the best estimates of such weights conditional on $\hat{\beta}$. If we look at these linear combinations as observable factors, this analysis delivers estimates of r $(< n)$ observable factors comprised of error-correction terms and lagged differences that can fully characterise the dynamics of Δy_t. An ordinary regression will produce the factor loadings.

Hecq et al. (2006) note that the common cycles hypothesis requires a very strict form of nested reduced rank structure in the VECM, namely that the left null space of Φ needs to be nested within that of Π, and hence r cannot be lower than q. They show that this makes the test for common cycles heavily dependent on the estimate of q, distorting the size of the test in finite samples. They relax this assumption and remove the nesting requirement from the common cycles hypothesis. That is, they study a system with reduced rank in Φ, where this rank is independent of the rank of Π. They call the resulting system a "weak form reduced rank structure." While the connection to the common BN cycles is lost, the weak form allows the rank of Φ to be anything between 1 to n, and this may result in more parsimonious models.

If the rank of Φ is r, then Φ can be written as $\gamma\delta'$ where γ and δ are $n \times r$ and $n(p-1) \times r$ matrices of full column rank. The reduced rank VECM becomes

$$\Delta y_t = c + \alpha\beta' y_{t-1} + \gamma\delta' w_{t-1} + \varepsilon_t, \tag{4}$$

and this has the nice interpretation that all dynamics in Δy_t are characterised by q linear indices made of lagged levels and r linear indices made of lagged differences. We call these indices "observable factors" instead of "indices" to distinguish them from those used in the "index models" defined by Sargent and Sims (1977). The determination of the rank of Φ is simply based on a partial canonical correlation analysis between Δy_t and w_{t-1} after the linear influence of z_{t-1} has been partialled out. One can then repeat the estimation of β, this time controlling for $\hat{\delta}' w_{t-1}$ instead of w_{t-1} and iterate between the two until convergence, although this adds extra procedural complexity to modelling, from which the gains are uncertain.

A VECM can be written in many ways, each with a different lag of y_t on the right hand side. In each of these isomorphic models, the coefficient of lagged y_t is Π, but the coefficient of w_{t-1} is a different linear combination of $(\pi_1, \ldots, \pi_p)$. If y_t share a small number of common cycles, all isomorphic ways that a VECM can be written have the same rank restrictions. However, if a VECM that has y_{t-1} on the right hand side has a weak form of reduced rank structure in which the rank of Φ is r, it may not have the same structure when it is rewritten using y_{t-p} instead. We follow Hecq et al. (2006) and restrict attention to reduced rank VECMs with y_{t-1} on the right hand side as in equation (4). According to this model, all dynamics in Δy_t can be fully explained by q factors made from the first lag of levels, and r factors made from lag 1 to lag $p-1$ of differences.

2.1 Sequential testing and the use of information criteria

We know that in VAR modelling the sequential reduction of individual parameters based on their individual significance can lead to models that deliver very bad forecasts. This might make a forecaster worried about using the sequential testing procedures discussed above for developing VAR models with strong or weak reduced rank structures. However, there is an important difference between the elimination of individual lags in each equation based on t-tests and tests of the ranks of the parameter matrices Π and Φ. Given the correlation among variables and correlation between adjacent lags, there is little information in a finite sample about the significance of a single lag of a single variable after the influence of all other regressors have been controlled for. That is why there are too many parameters with small t-values in a VAR, and final models can be different, depending on the order in which insignificant regressors have been eliminated.

In contrast, a test for a strong form reduced rank structure (i.e. for common cycles) looks for a linear combination of Δy_t that is unpredictable from the past, i.e. it is a test for a common statistical property or a "common feature" in the sense of Engle and Kozicki (1993). Each element of Δy_t possesses this property, but it is absent from a linear combination of Δy_t. While the data might have little power to determine the significance of the marginal contribution of individual lags, it is likely to have much more power in determining the joint significance of all lags. Also, since canonical variates are uncorrelated with each other by construction, reducing the rank of $\left(\Pi \vdots \Phi\right)$ by one neither affects the estimates nor the significance of the remaining indices made from $\left(z_{t-1}', w_{t-1}'\right)'$.

Similarly, in the reduced rank structures in (4), we are testing for the existence of a linear combination of Δy_t that, after controlling for z_{t-1}, cannot be predicted using lagged differences. In both forms, we start by determining the cointegrating rank, which searches for linear combinations of y_t that do not possess a common statistical property (the stochastic trend). The power of this procedure comes from the reduction in the stochastic order of magnitude of cointegrating combinations as compared with all other linear combinations of y_t. Although controlling for lagged differences is important for correct inference at this stage, the rank restrictions on Φ have no asymptotic influence on the results of the cointegration analysis. Also, since the cointegrating vectors are estimated superconsistently, conditional on q, the fact that they are estimated has no asymptotic influence on the second stage analysis. In practice, finite sample biases might affect the second stage analysis, but there are no theoretical results on the magnitude and direction of such effects.

Although we only use sequential tests in the empirical applications in this chapter, there have been some advances in developing appropriate information criteria for selecting p (the lag length), q (the cointegrating rank) and r (the rank of Φ) simultaneously. In the context of stationary VARs, Vahid and Issler (2002) examine the performance of modified versions of AIC, HQ and BIC in selecting p and r simultaneously. These criteria combine the log-likelihood value as a measure of fit with a penalty that counts the number of free parameters in each model. They find that the consideration of r as a choice variable improves the performance of BIC for sample sizes that are relevant in practice. In general they find that the HQ criterion is most successful in choosing the correct p and r.

Even though HQ and BIC can deliver consistent estimates of q as well as p and r in theory (Gonzalo and Pitarakis (1995) and Aznar and Salvador (2002)), their finite sample performance is very unreliable when determining q. Chao and

Phillips (1999) argue that the penalty term for an additional trending variable in a model has to be larger than that of a stationary variable. This means that the appropriate penalty depends on the true value of q which is not known. Based on this, they suggest using the "posterior information criterion (PIC)", which uses a data dependent penalty that implicitly accounts for the nature of trends in the data, to select p and q simultaneously. Athanasopoulos et al. (2009) consider using information criteria with a data dependent penalty similar to the PIC for simultaneous choice of p, q and r in reduced rank VECMs. While they find that such criteria perform well in choosing the correct model relative to other strategies, the derivation of the penalty function and coding it for each possible combination of (p, q, r) is challenging and adds to the procedural complexity of the modelling strategy. To simplify this, they consider a hybrid procedure that chooses p and r using the usual model selection criteria after removing the influence of y_{t-1} from Δy_t and w_{t-1}, and then using a data dependent penalty to choose q. This reduces the procedural complexity of the model selection stage without worsening the outcome.

Corander and Villani (2004) show that the fractional marginal likelihood criterion of O'Hagan (1995) can be used for consistent estimation of lag length and cointegrating rank. There has not been an extensive study of the finite sample performance of this criterion, but like HQ and BIC, it imposes the same penalty for an additional trending variable as for an extra stationary variable, so we expect that it will have the same difficulty in selecting q in finite samples.

2.2 The effects of cointegration on forecasting

Since economic theory often implies common stochastic trends in economic variables and these trends dominate the long-run forecasts of I(1) variables, the effects of cointegration on forecasting have been actively researched in the macroeconomic forecasting literature. See among others Engle and Yoo (1987), LeSage (1990), Clements and Hendry (1995), Lin and Tsay (1996), Hoffman and Rasche (1996), Christoffersen and Diebold (1998), Silverstovs et al. (2004), and Shoesmith (1995). This literature has tended to focus on the implication of the Granger representation theorem that VECMs are appropriate dynamic models for cointegrated variables. Unrestricted VAR in levels (UVARs) do not incorporate cointegrating restrictions and are therefore inefficient, while VARs in differences (DVARs) omit the important error-correction terms and are therefore under-specified. These observations have led to work that attempts to assess the gains from using VECMs for forecasting, and to assess a belief that for cointegrated I(1) variables, VECMs are likely to produce the best forecasts, followed by UVARs and then DVARs, especially for long-horizon forecasting. The results have not been clear-cut, even in simulation studies in which the nature of the trend in each variable, the lag length and the number of cointegrating vectors were known. Elliott (2006) provides a useful discussion on reasons for why this might be so. He uses a very simple stylised bivariate model to show that whether or not the use of error-correction terms improves short-run or long-run forecasting depends on almost all parameters in the design, including the covariance matrix of the errors.

2.3 Common nonlinear features and co-breaking

There are other forms of common features that imply cross equation restrictions in a multivariate autoregressive model. Anderson and Vahid (1998) define common nonlinear features, which occur when each variable in a multivariate time series has nonlinear dynamics, but there are linear combinations of these variables that have linear dynamics. These authors develop tests for common nonlinearity in the context of smooth transition VARs and VECMs, and they build a small empirical real business cycle model of the USA that incorporates common nonlinearity.

Co–breaking provides an example of a common nonlinear feature that has special relevance for forecasting. There is ample evidence that the structure of the data generating processes of many macroeconomic time series change over time, and these changes, especially level shifts, often provide the main reason for forecast failure in macroeconomic models (Clements and Hendry (2002)). One might expect that if such shifts are common to several variables, then the joint modelling of them would improve estimation precision and hence lead to improvements in forecasting. The situation in which shifts are common to a set of variables is called co-breaking. Clements and Hendry (1999) provide a detailed discussion of co-breaking and its possible presence in VECMs, and Hendry and Massmann (2007) review some more recent developments in this field. Most of the existing literature concentrates on how to incorporate common level shifts in time series models and conditional on knowing the times of such shifts, how to test if they are common. Questions such as how prevalent *common* level shifts are in economic variables, how they can be distinguished from stochastic trends, how best they can be modelled for the purpose of forecasting and how the concept of co-breaking can be generalised to common but non-contemporaneous shifts are important and challenging areas of research.

2.4 Other forms of parameter instabilities

As pointed out in the previous subsection, structural change is often present when working with macroeconomic data. One way that forecasters deal with the possibility of structural change is to use forecasting models that are developed using a fixed number of the most recent observations (a rolling window of fixed size). While rolling sample estimates can be helpful in models with only a few parameters, they are not necessarily helpful when working with VARs. In contrast to an expanding window that uses all available history, a rolling window does not improve the ratio of the sample size to the number of estimated parameters. In particular, the quality of the analysis of common trends does not improve and will always be based on a short span of data. In addition, the question of the optimal window size in the presence of structural breaks is a complex problem (see Pesaran and Timmermann (2007)).

Considering the problem of allowing for structural instability in a large scale latent factor model, Stock and Watson (2008) suggest that if the factors are estimated well, certain forms of structural instability can be accommodated by allowing time variation in just the factor loadings. Considering the reduced rank VECM in equation (4) as an observable factor model, one can consider the time varying model

$$\Delta y_t = c_t + \alpha_t \beta' y_{t-1} + \gamma_t \delta' w_{t-1} + \varepsilon_t.$$

This model can be justified on the basis that regime changes are temporary departures from fundamentals (i.e. temporary reductions in the strength of the error correcting mechanism and/or cycles) rather than permanent shifts in relationships between trends and between cycles in different variables. In the empirical example below, we consider a naive version of this in which q and r and the estimates of β and δ are determined from an expanding window, but the factor loadings and the intercept are estimated using a rolling window.

2.5 Shrinkage and Bayesian reduced rank VECMs

LeSage (1990) and Shoesmith (1995) demonstrate that shrinking the parameters of an error-correction model to zero (i.e. shrinking the model towards a random walk) proves to be beneficial in their forecasting applications. Canonical correlations analysis produces estimates of observable factors $\beta' y_{t-1}$ and $\delta' w_{t-1}$ that, after controlling for other factors, have equal variances and are orthogonal to each other. In addition, they are ranked on the basis of their partial contribution to the explanation of the variation in Δy_t. This provides a perfect setting for shrinkage towards models with smaller ranks. This can be achieved by shrinking the factor loadings towards zero, shrinking the less important factors more strongly to zero, or shrinking based on the L_1 norm as in Tibshirani (1996), which may result in some of the loadings to be set to zero.

There have been significant advances on the question of the choice of the shrinkage factor (or the tightness of the prior), but its discussion and application is beyond the scope of this Chapter. Here, we want to re-iterate the message in Shoesmith (1995) that VECMs and Bayesian VARs are not substitutes and there is scope for combining some aspects of both to produce better forecasting models. Moreover, the "raw" estimates[2] of β and δ from the canonical correlations proce-

[2] "Raw" estimates refer to the estimates that are produced under the normalisation that canonical covariates have unit sample variance and are uncorrelated with each other. Since any linear combination of cointegrating vectors is another cointegrating vector, it is customary to re-normalise the raw estimates such that a submatrix of the cointegrating vectors becomes a q dimensional identity matrix. This renormalisation can make the variances of each cointegrating

dures are already in a form that are ideal for the application of shrinkage to their loadings.

3 A Monte Carlo Illustration

We undertake a small Monte Carlo study to illustrate some of the issues concerning forecasting in reduced rank VECMs, and also to demonstrate the difficulty of deriving general methodological recommendations from Monte Carlo analysis. All analysis is based on one of the data generating processes (Model 4) originally used in Lin and Tsay (1996). This DGP for a four variable vector y_t is a VAR(2) in levels, specified by

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \varepsilon_t, \quad \text{with} \tag{5}$$

$$\pi_1 = \begin{pmatrix} 0.8 & 0.2 & 0 & 0.2 \\ .33 & -0.8 & .33 & -1.20 \\ 0.2 & -0.2 & 1 & -.2 \\ -.33 & 1.8 & -.33 & 2.20 \end{pmatrix}, \quad \pi_2 = \begin{pmatrix} -.15 & 0.15 & 0 & 0.15 \\ .3 & -0.09 & .3 & -0.06 \\ 0.15 & -0.15 & 0 & -.15 \\ -.3 & 0.09 & -.3 & 0.06 \end{pmatrix}$$

and $\varepsilon_t \sim IN(0_{4\times1}, I_{4\times4})$, and it has an equivalent error-correction representation

combination very different and destroy their uncorrelatedness.

which is

$$\Delta y_t = \Pi y_{t-1} + \Phi \Delta y_{t-1} + \varepsilon_t, \text{ with}$$

$$\Pi = \begin{pmatrix} -0.35 & 0.35 & 0 & 0.35 \\ .63 & -1.89 & .63 & -1.26 \\ 0.35 & -0.35 & 0 & -.35 \\ -.63 & 1.89 & -.63 & 1.26 \end{pmatrix} \text{ and } \Phi = -\pi_2.$$

The matrix Π has a rank of two, which implies that this DGP is driven by two independent unit root processes, and that there are two cointegrating vectors. Further, the rank of Φ is two, and it is clear that the null space of Π and Φ are the same. Hence y_t has two common cycles as well as two common trends. Study of the effects that restricted VAR modeling can have on forecast performance is particularly interesting when the true model is (1), because $p = q = r = 2$ in this case, providing scope for under and over specification of the ranks of Π and Φ, as well as scope for under and over specification of the model's lag length. While this DGP seems ideal for examining the effects of overestimating and underestimating p, q, and r for forecasting and we address these issues below, we also use it to illustrate that it is almost impossible to derive any general "rules of thumb" for forecasting.

We generate 10,000 series of samples of 120, 220 and 420 from equation (1) and use the first 100 (200 or 400) observations in each of the three samples for fixed window estimation, and the last twenty observations to assess forecast performance. Estimation is done via the procedure described in Section 2. Forecast performance is based on h-step ahead forecasts for $h = 1, 2, \ldots, 20$, and our forecast measures include the generalized forecast error second moment (GFESM) measure proposed by Clements and Hendry (1993), as well as the more traditional measures provided by the determinant and trace of the mean squared forecast error (MSFE) matrix. The GFESM is given by

$$GFESM_h = \det[E(e'_{T+1},\ e'_{T+2},\ \ldots, e'_{T+h})'(e'_{T+1},\ e'_{T+2},\ \ldots, e'_{T+h})]$$

where e_{T+j} is the j-step ahead forecast error, and like $det(MSFE_h)$, $GFESM_h$ has the attractive property that it remains invariant to elementary operations that involve different variables. $GFESM_h$ is the only one of these three measures that is invariant to elementary operations that involve the same variable at different horizons. A related property of $GFESM_h$ is that it provides a measure of forecast performance that relates to all forecasts up to and including $T + h$, whereas $det(MSFE_h)$ and $trace(MSFE_h)$ relate to just a single h-step ahead forecast. We take the h-th root of $GFESM_h$ to facilitate presentation with the other two measures.

3.1 Forecast performance of pre-specified models

Our Monte Carlo is based on Lin and Tsay's (1996) study, but we include an additional dimension in that we consider the forecasting consequences of the rank of Φ as well as the rank of Π. Imposing a VAR(2) specification (i.e. $p = 2$), we examine the implications for forecast performance, as we vary the cointegration rank in estimated models from $q = 0$ to $q = 4$, and vary the rank of Φ in the estimated model from $r = 0$ to $r = 4$. The models contain more free parameters (and are therefore less restricted) as q and r increase[3]. The special case in which $q = r = 0$ corresponds to a process containing four independent random walks. Cases in which $q = 0$ correspond to VAR(1) models in differences, whereas cases in which $q = 4$ correspond to VAR(2) models in levels. Short run dynamics are controlled by r, with $r = 0$ corresponding to a VAR(1) and $r = 4$ corresponding to the absence of any common cycles. We report forecast performances relative to the (estimated) benchmark set by the true specification in which $q = r = 2$, with a measure of 1.00 indicating a forecast performance that is as good as that produced by the true specification. An increase in any of the relative forecast measures indicates a deterioration in forecast performance.

The forecast results for this experiment are reported in the top two panels of

[3] This increase in free parameters is not linear, because the decrease in restrictions associated with an increase in the rank of a matrix is less than the increase in restrictions imposed as rank decreases.

Table 1, and to orientate the reader we point out that the centre of each block of twenty five forecasts reports a forecast performance measure of 1.00, which corresponds to the baseline model in which $q = r = 2$. Within each block, movement away from the centre corresponds to misspecification, and not surprisingly, forecast performance deteriorates in almost all cases. Random walk models nearly always produce the worst forecasts, regardless of the forecast horizon and the forecast measure.

Our $trace(MSFE_h)$ results for $r = 4$ correspond to those in Table III in Lin and Tsay (1996), and both sets of trace based results suggest that (i) the allowance for too much cointegration (too few unit roots) is detrimental for long horizon forecasts; and (ii) the imposition of too many unit roots (not enough cointegration) is detrimental for short horizon forecasts. The first point to make is that these conclusions are entirely dependent on the forecast loss function, and (i) is reversed if one looks at $Det(MSFE_h)$ or $GFESM_h$. Clements and Hendry (1995) make the same point based on their analysis of the two variable DGP originally analyzed by Engle and Yoo (1987).

In our setting, all forecast accuracy measures suggest that underspecification of cointegration is more damaging than overspecification for short horizon forecasts. However, we cannot claim that this a general result because the error correcting mechanism in this DGP is "strong", as evidenced by the non-zero eigenvalues for

Π being 0.35 and 0.63. Therefore, getting the cointegration structure right is quite important, and assuming a rank of zero or one is very costly. To confirm this we ran the same simulations but divided Π by 10 to make the error correcting mechanism weaker. Once we had done this, underspecification of the cointegrating rank had similar effects to overspecification, and very similar effects to the underspecification or overspecification of the rank of Φ (see the third and fourth panels of Table 1). Note that for the long horizon, even the trace measure now shows that severe underspecification of q is costlier than overspecification.

The variation in forecast performance associated with the variation of r is more subtle, but can be non-trivial when forecasting over short horizons. The greater importance of r for short horizon forecasts is not surprising given that the Φ matrix controls the short-run dynamics of the system. Further, it is intuitive that r should become less important when making long-run forecasts. In this example, the weaker forecast consequences of changing r relative to q might be attributable to a rather "weak" Φ matrix, (the non-zero eigenvalues are 0.03 and 0.15). This is supported by the results in the third and fourth panels of Table 1, which relate to a DGP for which eigenvalues of the parameter matrices $\frac{\Pi}{10}$ and Φ are similar. As with results associated with changing q, the severe underspecification of r is usually more damaging than overspecification of r, and this is easily rationalized once one recognizes that an underspecification of r, involves many more restrictions than an overspecification of r.

Overall, the Monte Carlo shows that for the DGP in (5), underspecification and overspecification of cointegration rank can have different consequences, depending on the loss function and forecast horizon. Misspecification of the rank of Φ has much milder effects. We emphasize that the patterns found here are data specific, and would change with the DGP. Elliott (2006) provides a good theoretical discussion on the imposition of cointegration when forecasting, and shows that the usefulness of this depends on the impact coefficient as well as serial correlation in the error-correction terms - both of which depend on the entire model (Π, Φ and the covariance of ε_t.). Misspecification of Φ did not have dire consequences in our case, perhaps because its weak non-zero eigenvalues had only weak effects on the system. On the other hand, underspecification of cointegration rank was quite problematic in our case, because of the "strength" of the cointegration implied by Π.

3.2 Model Selection and Forecast Performance

In practice, a modeler needs to choose the model lag length and the ranks of Π and Φ, and then conditioning on the forecast loss function and forecast horizon, forecast performance will depend on the choice of p, q and r, and how that choice influences forecasts relative to the true values of p, q and r. We conclude our Monte

Carlo exercise by modelling each simulated series, rather than specifying p, q and r in advance. Here, we separately use standard model selection criteria (AIC, HQ and BIC) to pick p, conduct the Johansen (1988) trace tests conditional on p to pick q (using the 5% critical values tabulated by MacKinnon et al. (1999)), and then conduct canonical correlations tests for rank reduction in Φ (at the 5% level) to pick r. We might expect our chosen lag structure to be relatively unimportant in this context, because the short-run dynamics in our system are not strong. Further we expect the cointegration testing stage to "work", firstly because we have true unit roots in the system, and secondly because the structure in our Π matrix is "strong".

We do not supply the details regarding model choices, but note that although AIC, HQ and BIC have very mixed success rates (89%, 54% and 5% respectively) in picking the true lag length, the success rate for picking the correct cointegration rank is always high (83%, 88% and 93%), regardless of the chosen p. The forecast performance measures are reported in Table 2, and these indicate that the chosen VECMs outperforms the UVARs for nearly all forecast measures and all forecast horizons, with the only exceptions occurring when AIC has been used to pick p and $GFESM$ has been used to assess forecasts for $h > 10$.

The tests for reduced rank in Φ are often inapplicable when HQ or BIC have been used to set p, because when $p = 1$, the $VECM(p-1)$ has no differenced lags. However, when these tests are applicable and observable difference factors are found and employed, the resulting RVECMs always offer slight improvements in forecasting relative to the VECMs. In our case the RVECM model is correct, so that improvements relative to VECMs are not surprising. Further, the use of RVECMs is particularly helpful for short run forecasting when AIC has been chosen to set the initial lag length - presumably because the reduced-rank restrictions offer another route to parsimony when the lag length is high.

We make two points about these results. First, for this DGP, BIC severely underestimates the lag length and hence it selects models that produce very inaccurate short-horizon forecasts. This provides a warning against the "rule of thumb" that models selected by BIC always produce better forecasts. Second, when the error-correcting mechanism is strong, the consequences of a wrong choice of q are severe, but the probability of a wrong choice of q is quite low.

4 Forecasting actual data

This section looks at an updated version of the well known Litterman (1986) data set, and discusses the use of vector autoregressions with reduced-rank coefficient matrices for forecasting in this context. The Litterman data set has often been used for forecast comparisons, and Shoesmith (1995) used it to compare the forecasting ability of VECMs, Bayesian VARs (BVARs), Bayesian VECMs (BECMs) and various other specifications. BECMs fared well in this context, leading Shoesmith to suggest that the use of Bayesian priors is slightly better for achieving parsimony than shorter lag structures. Here we compare the forecasting abilities of reduced rank models with unrestricted models.

Litterman's original model was based on a six variable VAR(6) that included quarterly data on seasonally adjusted real GNP, the implicit GNP price deflator, the unemployment rate, real gross private investment, the three month Treasury bill rate and the money supply. Our updated version of this data is drawn from FRED (Federal Reserve Economic Data) available from the Federal Reserve Bank of St Louis' website[4], and we convert the real GNP, implicit price deflator, investment and money supply series into natural logarithms. Our data consists of 203 observations from 1959q1 to 2009q3, and we initially use the first 91 observations (1959q1 to 1981q3) for estimation, leaving 112 observations to allow for the construction of 100 sets of 1 to 12 step ahead out of sample forecasts.

We consider the use of UVARs (with a maximum p of 6) and associated VECMs and RVECMs, and we use the model specification procedure and two step estimation procedures described earlier to construct both expanding and rolling window forecasts. We use the same measures of forecast performance as before, but now

[4]The series are GNPC96 (billions of chained 2005 $), GNPDEF (based on 2005), UNRATE (% of population16 and over), GPDIC96 (billions of chained 2005), TB3MS (annual %) and M1SL (billions of $). The original observations on UNRATE, TB3MS and M1SL are monthly, and they are converted to quarterly data by averaging.

present results relative to a benchmark expanding window UVAR with a lag length chosen by BIC.

The unit root characteristics of the initial estimation sample appear to be strong, in that unit root tests fail to reject the null for each series (even after significance levels are set to be >20%)[5]. Drift terms appear to be important in all but the unemployment and interest rate series. AIC, HQ and BIC respectively choose 4, 2 and 1 lags in the initial estimation period, and cointegration analysis on corresponding VECMS with 3, 1 and 0 lags in differences support two cointegrating relationships in the first two cases and three cointegrating relationships in the third. The implied error-correction terms exhibit strong first order serial correlation. Tests of reduced rank in Φ based on VARs in differences (with the lagged levels effects removed), respectively find three and two common cycles in the first two cases. In the third case we note that the Φ matrix is zero. The expanding window forecast performances are presented in Table 3. We discuss the first three panels first and then discuss the last panel separately.

The first three panels in Table 3 indicate that the standard VECM models consistently offer improvements over unrestricted VARs, and many of these improvements appear to be substantial We see, for instance a 40% improvement when using BIC to pick lag length and using $Trace(MSFE)$ to assess 12 step

[5]Shoesmith used inflation rather than log prices in his analysis, but unit root tests did not find strong support for this in our data set.

ahead forecasts, and a $(1 - .12/.58) = 80\%$ improvement when using models chosen by HQ and using $Det(MSFE)$ to assess 4-step-ahead forecasts. The percent reduction in forecast loss as we move from a UVAR to a VECM becomes more pronounced when $Det(MSFE)$ is used to assess forecasts, but this pattern is not as striking when looking at reductions as measured by $Trace(MSFE)$[6]. We often see small improvements when the VECM is restricted to a RVECM, but occasionally the VECM outperforms the RVECM (see the one step ahead $(GFESM)^{1/h}$ results for models with a lag structure chosen by AIC). Comparing models across each row in the table we see that the RVECM models based on lag structures chosen by HQ do best, and the UVAR with lag structures chosen by AIC do worst. The message that we take away from our forecast analysis of this data set is that extraneous parameterisation by AIC is very costly for forecasting. At the same time severe underparameterisation can be costly as well. The entire difference between HQ and BIC is due to a few initial periods in which BIC selects zero lags in differences while HQ selects a lag of one.

The third panel in Table 3 works with a loss function that emphasize unemployment and interest rates, because these variables are measured in percentages whereas the remaining variables have been converted to natural logarithms, lowering the relative contribution that these latter variables make to the $Trace(MSFE)$.

[6] We cannot make this comparison based on $(GFESM)^{1/h}$ because this measure is nonlinear in h.

Researchers often multiply the logarithms of variables by 100 so that differenced logs measure approximate percentage change, and it is instructive to see how this can change the relative ranking of forecasts. The scaling won't change the selected model (apart from scaling of coefficients and standard errors), but portions of the loss function are now exaggerated due to the higher standard errors of the associated regression equations, so that patterns in the relative ranking of models according to $Trace(MSFE)$ can change, whereas the rankings according to the other two measures are invariant. The final panel in Table 3 shows what happens in this case. Before scaling, the BIC chosen VECM and RVECM models performed well relative to UVARs when forecasting over longer horizons, according to the third panel. However after scaling, the UVAR, VECM and RVECM's deliver very similar longer horizon forecasts according to the $Trace(MSFE)$ measure reported in the fourth panel. A comparison of performance ranking in the third and fourth panels therefore illustrates the scaling problems associated with using $Trace(MSFE)$.

In contrast to the Monte Carlo study presented earlier, the real data studied here is likely to contain near unit roots and trends as well as undergo structural change. This makes it harder for accurate model selection to occur, and harder to forecast, regardless of the forecast horizon and loss function. We close our empirical example with two further exercises. The first of these calculates and compares rolling window forecast with the expanding window forecasts, while the second follows the ideas in Stock and Watson (2008), and is based on expanding window estimates of error-correction terms and observable difference factors, but the adjustment and factor loading coefficients are estimated using only the last eighty observations. We call these models TVFL models.

The performances of the rolling window forecasts are presented in Table 4, and for each line in the table, the baseline is the reference benchmark used in the expanding window analysis - i.e. the UVAR chosen by BIC. The first three columns for each information criterion correspond to those in Table 3, and the patterns noted above in Table 3 are present in this table as well. In particular, if we pick the best and worst performer in each line, then the winners and losers in Table 4 almost perfectly correspond with those in Table 3. An important conclusion arising from the results in the two tables is that the rolling window forecasts are no better than expanding window forecasts. For each information criterion, the forecasts from the TVFL models are typically better than those from the standard RVECM rolling window forecasts, but they are rarely better than the corresponding RVECM expanding window forecasts. We conclude that naive strategies for safe-guarding against structural change do not improve forecast performance in this data set.

5 Conclusion

In this chapter we present VAR models with cointegration and common cycles (or weaker forms of rank restrictions), and show that they can be usefully viewed as observable factor models. The factors are linear combinations of lagged levels and lagged differences, and as such, they offer potential for forecasting.

In our empirical application we observe that as long as excessively parameterised models are avoided, the use of rank restrictions improve forecasts. Our small Monte Carlo analysis illustrates that the question of whether rank restrictions are important for forecasting depends on many aspects of the DGP. This makes it virtually impossible to develop "one size fits all" prescriptions for practitioners. However, if there are compelling reasons to consider common trends or common cycles in a set of variables, as it is for a set of macroeconomic aggregates, then one should consider reduced rank error-correction models as possible forecasting tools.

Acknowledgments:

We thank the Australian Research Council for financial support (under grant DP0984399), Taya Dumrongrittikul, Yin Liao, Shuping Shi and Wenying Yao for research assistance, and Michael Clements and an anonymous referee for helpful comments. We are responsible for all remaining errors and omissions.

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.

Akaike, H. (1976). Canonical correlation analysis of time series and the use of information criterion. In R. Mehra and D. Lianiotis (Eds.), *System Identification.* New York: Academic Press.

Anderson, H. M., C. N. Low, and R. Snyder (2006). Single source of error state space approach to the Beveridge-Nelson decomposition. *Economic Letters 91*, 104–109.

Anderson, H. M. and F. Vahid (1998). Testing multiple equation systems for common nonlinear components. *Journal of Econometrics 84*, 1–36.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics 22*, 327–351.

Athanasopoulos, G., O. T. Guillen, J. V. Issler, and F. Vahid (2009). Model selection, estimation and forecasting in VAR models with short-run and long-run restrictions. *Working Paper, Department of Econometrics and Business Statistics, Monash University 2/09.*

Aznar, A. and M. Salvador (2002). Selecting the rank of the cointegration space and the form of the intercept using an information criterion. *Econometric Theory 18*, 926 – 947.

Beveridge, S. and C. R. Nelson (1981). A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics 7*, 151–174.

Box, G. E. P. and G. C. Tiao (1977). A canonical analysis of multiple time series. *Biometrika 64*, 355–365.

Chao, J. C. and P. C. B. Phillips (1999). Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics 91*, 227–271.

Christoffersen, P. F. and F. X. Diebold (1998). Cointegration and long-horizon forecasting. *Journal of Business and Economic Statistics 16*, 450–458.

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

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 (1999). *Forecasting non-stationary economic time series*. MIT Press.

Clements, M. P. and D. F. Hendry (2002). Modelling methodology and forecast failure. *Econometrics Journal 5*, 319–344.

Corander, J. and M. Villani (2004). Bayesian assessment of dimensionality in multivariate reduced rank regression. *Statistica Neerlandica 50*, 255 – 270.

Doan, T., R. B. Litterman, and C. A. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*, 1–10.

Elliott, G. (2006). Forecasting with trending data. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1. North-Holland.

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

Engle, R. F. and S. Kozicki (1993). Testing for common features. *Journal of Business and Economic Statistics 11*, 369–395.

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

Gonzalo, J. and J.-Y. Pitarakis (1995). Specification via model selection in vector error correction models. *Economic Letters 60*, 321 – 328.

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

Hecq, A., F. Palm, and J.-P. Urbain (2006). Common cyclical features analysis in VAR models with cointegration. *Journal of Econometrics 132*, 117–141.

Hendry, D. F. and M. Massmann (2007). Co-Breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics 25*, 33 – 51.

Hoffman, D. and R. Rasche (1996). Assessing forecast performance in a cointegrated system. *Journal of Applied Econometrics 11*, 495–517.

Johansen, S. (1988). Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control 12*, 231–254.

Johansen, S. (1995). *Likelihood-based Inference on Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Kadiyala, K. R. and S. Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR models. *Journal of Applied Econometrics 12*, 99–132.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews 29*, 224 – 242.

LeSage, J. P. (1990). A comparison of the forecasting ability of ECM and VAR models. *Review of Economics and Statistics 72*, 664–671.

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

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business and Economic Statistics 4*, 25–38.

MacKinnon, J. G., A. A. Haug, and L. Michelis (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics 14*, 563–577.

O'Hagan, A. (1995). Fractional Bayes factors for model comparisons. *Journal of the Royal Statistical Society B 57*, 99 – 138.

Pesaran, M. H. and A. Timmermann (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics 137*, 134–161.

Reinsel, G. C. and S. K. Ahn (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio tests and forecasting. *Journal of Time Series Analysis 13*, 353–375.

Sargent, T. J. and C. A. Sims (1977). Business cycle modelling without pretending to have too much of prior economic theory. In C. Sims (Ed.), *New Methods in Business Cycle Research*, pp. 45–100. Federal Reserve Bank of Minneapolis.

Shoesmith, G. L. (1995). Multiple cointegrating vectors, error correction, and forecasting with Litterman's model. *International Journal of Forecasting 11*, 557–567.

Silverstovs, B., T. Engsted, and N. Haldrup (2004). Long-run forecasting in multicointegrated systems. *Journal of Forecasting 23*, 315–335.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*, 1–48.

Sims, C. A. and T. A. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review 39*, 949–968.

Stock, J. H. and M. W. Watson (2008). Forecasting in dynamic factor models subject to structural instability. In J. Castle and N. Shephard (Eds.), *The methodology and practice of econometrics, A Festschrift in Honour of Professor David F. Hendry*. Oxford University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society B 58*, 267–288.

Vahid, F. (2006). Common cycles. In K. D. Patterson and T. C. Mills (Eds.), *Pal-*

grave Handbook of Econometrics, Volume 1: Econometric Theory, Chapter 16, pp. 610–630. Basingstoke: Palgrave-McMillan.

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

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

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

Waggoner, D. and T. A. Zha (1999). Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics 81*, 639–651.

A Canonical correlations and partial canonical correlations

This appendix contains a very brief overview of canonical correlations analysis. Please refer to Hamilton (1994) for a more thorough exposition.

The R^2 of a regression is the largest squared sample correlation between a single dependent variable and all possible linear combination of independent variables. The squared sample canonical correlations between an $n \times 1$ vector y_t and a $k \times 1$

vector x_t based on a sample of T observations are the multivariate equivalents of R^2 for the case in which $n > 1$. The largest one measures the largest squared sample correlation between all possible linear combinations of y_t and all possible linear combinations of x_t. The second largest measures the largest squared correlation among all linear combinations that are orthogonal to the first combination, and so on. Obviously there will only be $\min\{n, k\}$ non-zero sample squared correlations. Squared canonical correlations are the eigenvalues of

$$(Y'Y)^{-1} Y'X (X'X)^{-1} X'Y \tag{6}$$

which are the same as the eigenvalues of the symmetric matrix

$$(Y'Y)^{-\frac{1}{2}} Y'X (X'X)^{-1} X'Y (Y'Y)^{-\frac{1}{2}}. \tag{7}$$

In the multivariate regression model

$$\underset{T\times n}{Y} = \underset{T\times k}{X}\underset{k\times n}{B} + \underset{T\times n}{E}$$

where $n < k$, testing the statistical significance of $n - r$ smallest squared canonical correlations is a test for the hypothesis that the rank of the parameter matrix B is at most r, because it tells us that there is no need for more than r linear combinations of x_t to explain the entire linear dependence between y_t and x_t.

Denoting the squared correlations by $\lambda_1 < \lambda_2 < \cdots < \lambda_n$, the maximum of the log-likelihood value of the model with rank r under the assumption of normality is (ignoring the constants that do not depend on r)

$$-\frac{T}{2}\ln|Y'Y| - \frac{T}{2}\sum_{i=n-r+1}^{n}\ln(1-\lambda_i)$$

This implies that the likelihood ratio test of the hypothesis that rank of B is at most r against the alternative that it is of full rank is simply

$$-T\sum_{i=1}^{n-r}\ln(1-\lambda_i).$$

The assumption of normality can be relaxed by analysing this problem in a generalised method of moments framework (see, for example, Anderson and Vahid (1998)). There are $(n-r)(k-r)$ restrictions involved in restricting rank of B from n to r. Hence under classical assumptions in a regression framework and stationarity and ergodicity of y_t and x_t in a time series framework the asymptotic distribution of this test statistic is $\chi^2_{(n-r)(k-r)}$.

A single calculation of eigenvalues of the matrix in (7) delivers all that is needed to determine rank of B through a sequence of tests. Moreover, if B has rank r and hence can be written as CD' where C and D are $n \times r$ and $k \times r$ matrices of rank r, the OLS regression of the combinations of Y made using the eigenvectors corresponding to r largest eigenvalues of the matrix (6) on X provides the maximum likelihood estimates of D. Then another OLS regression of Y on $\hat{D}'X$ provides the maximum likelihood estimates of C.

Partial canonical correlations are canonical correlations after the linear influence of a set of variables z_t is removed from both y_t and x_t. They can be used for testing of the rank of B in the multivariate regression

$$\underset{T\times n}{Y} = \underset{T\times k}{X}\underset{k\times n}{B} + \underset{T\times m}{Z}\underset{m\times n}{\Gamma} + \underset{T\times n}{E}$$

and when Γ is unrestricted. Similarly, one can obtain the maximum likelihood estimate of a reduced rank B from this analysis.

Table 1: Relative Forecast Performance of Different Specifications of the VAR(2)

	$(\text{GFESM})^{1/h}$					Det(MSFE)					Trace(MSFE)				
One step ahead forecasts															
q/r	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	16.65	2.98	2.46	2.53	2.54	16.65	2.98	2.46	2.53	2.54	4.21	1.42	1.35	1.35	1.36
1	1.71	1.33	1.14	1.17	1.17	1.71	1.33	1.14	1.17	1.17	1.17	1.09	1.04	1.05	1.05
2	1.24	1.01	1.00	1.01	1.02	1.24	1.01	1.00	1.01	1.02	1.07	1.00	1.00	1.00	1.00
3	1.31	1.06	1.05	1.07	1.07	1.31	1.06	1.05	1.07	1.07	1.08	1.01	1.01	1.02	1.02
4	1.32	1.07	1.06	1.08	1.08	1.32	1.07	1.06	1.08	1.08	1.08	1.02	1.01	1.02	1.02
Twenty step ahead forecasts															
q/r	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	1.42	1.30	1.25	1.25	1.25	5.16	3.29	2.22	2.21	2.20	1.07	1.00	1.00	1.00	1.00
1	1.16	1.11	1.08	1.08	1.08	2.89	2.31	1.85	1.85	1.85	1.00	1.00	1.00	1.00	1.00
2	1.03	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.00	1.00
3	1.03	1.00	1.01	1.01	1.01	1.18	1.21	1.23	1.24	1.24	1.11	1.12	1.13	1.13	1.13
4	1.03	1.00	1.00	1.00	1.00	1.16	1.19	1.22	1.24	1.24	1.14	1.15	1.16	1.17	1.17
One step ahead forecasts for the DGP with $\Pi/10$															
q/r	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	1.71	1.38	1.30	1.33	1.33	1.71	1.38	1.30	1.33	1.33	1.23	1.12	1.10	1.10	1.10
1	1.26	0.93	0.90	0.92	0.92	1.26	0.93	0.90	0.92	0.92	1.08	0.98	0.98	0.98	0.98
2	1.37	1.02	1.00	1.02	1.02	1.37	1.02	1.00	1.02	1.02	1.10	1.01	1.00	1.00	1.00
3	1.42	1.07	1.05	1.07	1.07	1.42	1.07	1.05	1.07	1.07	1.11	1.02	1.01	1.02	1.02
4	1.42	1.07	1.05	1.07	1.08	1.42	1.07	1.05	1.07	1.08	1.11	1.02	1.01	1.02	1.02
Twenty step ahead forecasts for the DGP with $\Pi/10$															
q/r	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	1.07	1.07	1.06	1.07	1.07	2.21	2.13	2.10	2.10	2.10	1.43	1.39	1.39	1.39	1.39
1	1.04	0.99	0.98	0.99	0.99	1.39	0.88	0.83	0.82	0.82	1.06	0.93	0.93	0.93	0.93
2	1.04	1.00	1.00	1.00	1.00	1.31	1.02	1.00	1.01	1.01	1.07	1.00	1.00	1.00	1.01
3	1.05	1.01	1.02	1.02	1.02	1.48	1.28	1.30	1.32	1.33	1.14	1.11	1.11	1.12	1.12
4	1.05	1.01	1.02	1.02	1.02	1.39	1.24	1.28	1.31	1.31	1.18	1.16	1.17	1.18	1.18

Notes:
1. q and r are ranks of Π and Φ matrices in equation (3) in the text.
2. We report all forecast measures relative to the correct specification (q=r=2)
3. The reported results relate to 10000 simulations of estimation samples of 100.
4. Results for samples of 200 and 400 are qualitatively similar, but show less variation from the baseline.

Table 2: Relative Forecast Performance of Chosen Specifications

	$(\text{GFESM})^{1/h}$			Det(MSFE)			Trace(MSFE)		
One step ahead forecasts									
model type/lag selection	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC
UVAR	1.12	1.21	1.31	1.12	1.21	1.31	1.03	1.05	1.08
VECM	1.08	1.16	1.24	1.08	1.16	1.24	1.02	1.04	1.07
RVECM	1.04	1.13	1.24	1.04	1.13	1.24	1.01	1.03	1.07
Five step ahead forecasts									
model type/lag selection	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC
UVAR	1.06	1.08	1.11	1.27	1.23	1.16	1.13	1.11	1.08
VECM	1.05	1.06	1.07	1.15	1.08	0.99	1.02	1.01	0.99
RVECM	1.04	1.05	1.07	1.10	1.05	0.98	1.00	1.00	0.99
Ten step ahead forecasts									
model type/lag selection	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC
UVAR	1.03	1.05	1.06	1.35	1.30	1.25	1.20	1.18	1.15
VECM	1.04	1.04	1.05	1.14	1.08	1.00	1.02	1.01	1.00
RVECM	1.03	1.04	1.04	1.12	1.07	1.00	1.00	1.00	1.00
Twenty step ahead forecasts									
model type/lag selection	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC
UVAR	1.01	1.02	1.02	1.25	1.22	1.16	1.18	1.16	1.14
VECM	1.02	1.03	1.03	1.11	1.06	0.98	1.01	1.01	1.00
RVECM	1.02	1.02	1.03	1.11	1.06	0.98	1.00	1.00	1.00

Notes:

1. We report all forecast measures relative to the correct specification (p=q=r=2)
2. The reported results relate to 10000 simulations of estimation samples of 100.
3. UVAR models allow for full rank in Π and Φ
4. VECM models have reduced rank in Π, but allow for full rank in Φ
5. RRVECM models have reduced rank in Π and Φ
6. Results for samples of 200 and 400 reinforce the finding that RRVECMs dominate the VECMs, which dominate the UVARs.

Table 3: Relative Forecast Performance of Reduced Rank Models of the Litterman Data

	AIC			HQ			BIC		
h	UVAR	VECM	RVECM	UVAR	VECM	RVECM	UVAR	VECM	RVECM
				$(GFESM)^{1/h}$					
1	0.89	0.53	0.66	0.53	0.31	0.30	1.00	0.69	0.64
4	1.65	1.07	1.07	0.85	0.54	0.54	1.00	0.72	0.71
8	1.74	1.23	1.16	0.93	0.79	0.78	1.00	0.89	0.88
12	1.69	1.28	1.23	0.98	0.92	0.92	1.00	0.95	0.94
				$Det(MSFE)$					
1	0.89	0.53	0.66	0.53	0.31	0.30	1.00	0.69	0.64
4	0.72	0.20	0.20	0.58	0.12	0.12	1.00	0.32	0.30
8	1.65	0.32	0.33	0.72	0.22	0.22	1.00	0.52	0.50
12	3.30	0.59	0.58	0.88	0.37	0.36	1.00	0.58	0.56
				$Trace(MSFE)$					
1	1.09	1.01	1.15	1.04	0.95	0.94	1.00	0.96	0.93
4	1.26	1.14	1.20	0.97	0.65	0.66	1.00	0.65	0.65
8	1.28	1.07	1.11	0.99	0.56	0.56	1.00	0.60	0.60
12	1.36	1.10	1.10	1.00	0.57	0.57	1.00	0.60	0.60
				$Trace(MSFE) - 100 \times \{ln(GNP), ln(INV), ln(M1), ln(Price)\}$					
1	1.17	1.14	1.11	0.89	0.94	0.92	1.00	1.00	1.00
4	1.19	1.15	1.16	0.96	1.00	0.99	1.00	1.03	1.01
8	0.90	0.79	0.79	0.99	0.91	0.91	1.00	1.06	1.05
12	1.02	0.84	0.85	1.01	0.88	0.88	1.00	0.97	0.97

Notes:
1..Results are based on expanding window estimates and are relative to the UVAR chosen by BIC
2. See footnotes 3 - 5 for characteristics of UVARS, VECMs and RVECMs
3. In the last panel, $lnGDP$, LnP, $LnInv$ and LnM have been scaled by 100.

Table 4: Relative Forecast Performance of Reduced Rank Models of the Litterman Data

	AIC				HQ				BIC			
h	UVAR	VECM	RVECM	TVFL	UVAR	VECM	RVECM	TVFL	UVAR	VECM	RVECM	TVFL
						$(GFESM)^{1/h}$						
1	1.53	1.01	0.92	0.72	0.58	0.33	0.32	0.30	1.72	1.47	1.39	0.61
4	3.18	2.20	2.11	1.35	0.95	0.67	0.67	0.61	1.09	1.01	0.98	0.76
8	2.57	2.17	2.06	1.38	1.02	0.87	0.86	0.84	1.09	1.04	1.01	0.93
12	2.29	1.98	1.95	1.45	1.07	0.94	0.95	0.94	1.01	0.98	0.96	0.97
						$Det(MSFE)$						
1	1.53	1.01	0.92	0.72	0.58	0.33	0.32	0.30	1.72	1.47	1.39	0.61
4	1.25	0.51	0.45	0.27	0.49	0.12	0.13	0.15	1.26	0.96	0.90	0.32
8	1.37	0.52	0.51	0.50	0.67	0.20	0.21	0.31	1.30	1.16	1.15	0.56
12	2.13	0.69	0.69	0.95	0.73	0.28	0.29	0.59	1.23	1.18	1.18	0.79
						$Trace(MSFE)$						
1	1.37	1.26	1.29	1.28	1.17	1.08	1.06	1.00	1.17	1.07	1.05	0.99
4	1.60	1.18	1.16	1.35	1.10	0.70	0.73	0.74	1.17	0.85	0.85	0.72
8	1.41	0.95	0.94	1.28	1.12	0.54	0.55	0.68	1.13	0.70	0.70	0.72
12	1.28	0.89	0.86	1.26	1.10	0.51	0.52	0.69	1.08	0.66	0.66	0.72

Notes:

1..Results are based on rolling window estimates and are relative to the expanding window UVAR estimates chosen by BIC

2. TVFL models incorporate factor loading estimates that are derived from the last eighty observations in the rolling sample