IMPULSE RESPONSE AND FORECAST ERROR VARIANCE ASYMPTOTICS IN NONSTATIONARY VARS

BY

PETER C.B. PHILLIPS

COWLES FOUNDATION PAPER NO. 953

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
1998

ELSEVIER Journal of Econometrics 83 (1998) 21–56

JOURNAL OF Econometrics

Impulse response and forecast error variance asymptotics in nonstationary VARs

Peter C.B. Phillips

Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281, USA

(Received 1 May 1995; received in revised form 1 March 1996)

".. though the models forecast well over horizons of four to six quarters, they disagree so strongly about the effects of important monetary and fiscal policies that they cannot be considered reliable guides to such policy effects, until it can be determined which of them are wrong and which (if any) are right." Christ (1975, p. 54)

Abstract

Estimated impulse responses and forecast error decompositions are shown to be inconsistent at long horizons in unrestricted VARs with some unit roots. Predictions from unrestricted VARs also do not converge to the optimal predictors over long forecast horizons. In contrast, reduced rank regressions produce impulse responses and forecast error variance estimates that are consistent and predictions that are asymptotically optimal, provided the cointegrating rank is correctly specified or consistently estimated by an order selector such as PIC. Some simulations show these findings to be relevant in finite samples in VARs with some unit roots and cointegration. © 1998 Elsevier Science S.A.

Key words: Error correction model; Forecast error variance decomposition asymptotics; Impulse response asymptotics; Reduced rank regression; Vector autoregression; Unit-root asymptotics
JEL classification: C32

Presented at the Conference in Honor of Carl F. Christ, Johns Hopkins University, April 21–22 1995. The computations and graphics reported here were performed by the author on a P5-66 using programs written in GAUSS. The paper was keyboarded by the author in MSWord 6. Thanks go to two referees for comments on the original version and to the NSF for research support under grant nos. SES 91-22142 and SBR 94-22922.

PII S0304-4076(97)00064-X

1. Introduction

Twenty years ago Christ (1975) published a study that set out to judge the performance of several well-established structural econometric models of the US economy. Christ's study appeared as part of a general symposium on econometric model performance that was published in the *International Economic Review* during 1974–75. His study compared the post-sample, post-model building forecast performance of nine different models, using forecast root mean squared errors (RMSEs) to evaluate the results. In addition, Christ considered the multiplier effects over time of certain monetary and fiscal policy shocks to the models' exogenous variables. One of Christ's key discoveries was that there was great uncertainty across models about the macroeconomic effects that follow from important fiscal and monetary policy actions. For instance, in studying the effect of an easing of monetary policy (measured as an increase of $1 billion in unborrowed reserves) on real GNP, Christ found that serious disagreement among the models set in almost immediately after the policy change. Some models showed only positive effects, others showed positive and negative effects on real GNP over time; some gave monotonic effects, others showed cyclical effects; some seemed to converge, others to diverge. Against this background of disparity among leading econometric models of the US economy, Christ concluded that the models could not be relied upon as guides to the effects of economic policies even though their forecasting performance was quite respectable.

Since the 1970s there has been less reliance, at least in academic research, on large structural econometric models (of the type studied in Christ's paper) for policy analysis purposes. Instead, more attention has been given to small-scale time-series models like vector autoregressions (VARs) as instruments of policy analysis, and VARs and Bayesian VARs (BVARs) as tools of prediction. These models are often regarded by their users as having fewer subjective design elements than large structural econometric models. Nevertheless, they are far from being objective tools of prediction or policy analysis and, at least as they stand, they are certainly not automated modelling devices. Similar remarks apply to reduced rank regression (RRR) models and error correction models (ECMs), which are now becoming popular in the analysis of macroeconomic time series. These models come within the general category of VAR systems but explicitly incorporate certain information about the existence of unit roots and the presence of cointegrating relations among the variables. Such information can be either specified a priori or data-based (i.e. determined by the sample data). Either way, it constitutes a design feature that will certainly affect both forecasts and policy analysis.

The present paper studies how design features of the type just mentioned affect the large sample behaviour of VAR forecasts and policy analyses. Although they are certainly an important element in practical VAR modelling and do indeed enter into our simulation exercises, we will not be specifically concerned in this paper with model selection or model comparison issues. These have recently been extensively discussed in this context in other work by the author (1996). Instead, this paper seeks to develop an asymptotic theory for forecasting and policy analysis with VARs that allows for nonstationary elements (specifically, unit roots, near-unit roots and cointegration) and evaluates how design features in the models that accommodate these elements affect the asymptotic performance of these model characteristics. We are specifically interested in comparing unrestricted VARs with RRRs and ECMs in forecasting and policy analysis, because these models are the backbone of much ongoing empirical analysis and because they highlight the differences that are known to occur in the asymptotic theory of estimation of nonstationary VARs. In that theory, the role of prior information in the asymptotics is substantial and determines not only whether coefficient estimators are efficient but also whether they are asymptotically unbiased – see Phillips (1991) for details. This paper examines the role of prior information (or data-determined model selection) on asymptotic forecast performance and policy analysis by considering the limiting behavior of a system's impulse responses, its estimated forecast error variance matrices and their associated decompositions.

We start our analysis by showing that impulse responses that are calculated from unrestricted VARs with roots near unity have long period estimated impulse responses that are inconsistent. In fact, these estimated impulse responses tend to random variables rather than the true impulse responses as the sample size increases. Hence, policy analysis that is undertaken from unrestricted VARs using estimated impulse responses can be expected to be inherently uncertain even in large samples as the horizon increases. Models that explicitly determine the presence and number of unit roots like data-based RRRs (see Phillips, 1994, 1996), in which the cointegrating rank is consistently estimated, do not suffer from this difficulty asymptotically. However, these models as well as BVARs that are formulated with unit root priors all suffer from the same problem to a greater or lesser extent in finite samples.

Forecast error variance decompositions are also inconsistent in unrestricted VAR models with near unit roots. An interesting feature of this finding is that since the estimated prediction error variance of an unrestricted VAR with some roots near unity is a random variable in the limit, it turns out that there is an appreciable probability (0.68 in a random-walk model) that the estimated prediction error variance is less than the actual prediction error variance of the optimal predictor. This means that unrestricted VAR regressions give inconsistent estimates of the forecast error variance at long horizons, and also have a tendency to understate this variance.

We conduct simulation exercises to assess the sensitivity of forecasting performance and policy analysis to specific design features of models within the general VAR class. We look at short and long period ahead forecasts and multiplier effects, and we consider models with and without unit roots and cointegrating relations. Our general conclusion, like Christ's, is that, while there are some notable differences in forecasting performance, the biggest differences occur in policy analysis. Apparently, minor differences in models that seem to have little overall effect on average forecasting performance can have really substantial effects on policy analyses. This is especially true when there are unit roots or near unit roots in the fitted model.

2. Impulse response asymptotics with some roots at, or near, unity

In stationary VARs the system's estimated impulse responses and forecast error variance decompositions are $\sqrt{n}$-consistent and, upon appropriate centering and scaling, they have asymptotic normal distributions. The calculations leading to the limit theory are straightforward and simply make use of the functional representations of these quantities in terms of the estimated VAR coefficients, the limiting normal distribution of the latter and the continuous mapping theorem. Lutkepohl (1994, Chapter 3.7) provides derivations along these lines. When there are some unit roots in the VAR system, the limit theory of the estimated VAR coefficients changes and has some nonnormal components – see Phillips and Durlauf (1986), Park and Phillips (1988, 1989) and Sims et al. (1990). In this case the full matrix of estimated regression coefficients in a VAR is asymptotically normal but singular to the extent that there are some components in the system like unit roots and cointegrating vectors that converge at a faster rate. Since the estimated VAR coefficients are consistent (and, indeed, converge at faster rates in some directions) it might reasonably be expected that the impulse responses are also. However, as we show below, this is not the case for long horizon impulses. Moreover, since the estimated VAR coefficients have a limit normal distribution, albeit singular, it might also be anticipated that the impulse responses would be asymptotically normal. Again, we show this not to be the case. There are therefore some major differences in the limit theory of impulse responses between stationary and nonstationary VARs and these differences do seem to be important in the analysis of empirical results.

Let y_t be a m-vector time series generated by the following pth order VAR model

$$y_t = J(L)y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{1}$$

where $J(L) = \sum_{i=1}^{p} J_i L^{i-1}$. The system (1) is initialized at $t = -p+1, \ldots, 0$ and we may let these initial values be any random vectors including constants. It is often convenient to set the initial conditions so that the I(0) component of (1) is stationary and we will proceed as if this has been done. The presence of deterministic components in (1) does not affect our conclusions in any substantive way, so we will proceed as if they are absent just to keep the derivations as simple as possible. It is also convenient to write (1) in levels and differences this paper with model selection or model comparison issues. These have recently been extensively discussed in this context in other work by the author (1996). Instead, this paper seeks to develop an asymptotic theory for forecasting and policy analysis with VARs that allows for nonstationary elements (specifically, unit roots, near-unit roots and cointegration) and evaluates how design features in the models that accommodate these elements affect the asymptotic performance of these model characteristics. We are specifically interested in comparing unrestricted VARs with RRRs and ECMs in forecasting and policy analysis, because these models are the backbone of much ongoing empirical analysis and because they highlight the differences that are known to occur in the asymptotic theory of estimation of nonstationary VARs. In that theory, the role of prior information in the asymptotics is substantial and determines not only whether coefficient estimators are efficient but also whether they are asymptotically unbiased – see Phillips (1991) for details. This paper examines the role of prior information (or data-determined model selection) on asymptotic forecast performance and policy analysis by considering the limiting behavior of a system's impulse responses, its estimated forecast error variance matrices and their associated decompositions.

We start our analysis by showing that impulse responses that are calculated from unrestricted VARs with roots near unity have long period estimated impulse responses that are inconsistent. In fact, these estimated impulse responses tend to random variables rather than the true impulse responses as the sample size increases. Hence, policy analysis that is undertaken from unrestricted VARs using estimated impulse responses can be expected to be inherently uncertain even in large samples as the horizon increases. Models that explicitly determine the presence and number of unit roots like data-based RRRs (see Phillips, 1994, 1996), in which the cointegrating rank is consistently estimated, do not suffer from this difficulty asymptotically. However, these models as well as BVARs that are formulated with unit root priors all suffer from the same problem to a greater or lesser extent in finite samples.

Forecast error variance decompositions are also inconsistent in unrestricted VAR models with near unit roots. An interesting feature of this finding is that since the estimated prediction error variance of an unrestricted VAR with some roots near unity is a random variable in the limit, it turns out that there is an appreciable probability (0.68 in a random-walk model) that the estimated prediction error variance is less than the actual prediction error variance of the optimal predictor. This means that unrestricted VAR regressions give inconsistent estimates of the forecast error variance at long horizons, and also have a tendency to understate this variance.

We conduct simulation exercises to assess the sensitivity of forecasting performance and policy analysis to specific design features of models within the general VAR class. We look at short and long period ahead forecasts and multiplier effects, and we consider models with and without unit roots and cointegrating relations. Our general conclusion, like Christ's, is that, while there are some notable differences in forecasting performance, the biggest differences occur in policy analysis. Apparently, minor differences in models that seem to have little overall effect on average forecasting performance can have really substantial effects on policy analyses. This is especially true when there are unit roots or near unit roots in the fitted model.

2. Impulse response asymptotics with some roots at, or near, unity

In stationary VARs the system's estimated impulse responses and forecast error variance decompositions are $\sqrt{n}$-consistent and, upon appropriate centering and scaling, they have asymptotic normal distributions. The calculations leading to the limit theory are straightforward and simply make use of the functional representations of these quantities in terms of the estimated VAR coefficients, the limiting normal distribution of the latter and the continuous mapping theorem. Lutkepohl (1994, Chapter 3.7) provides derivations along these lines. When there are some unit roots in the VAR system, the limit theory of the estimated VAR coefficients changes and has some nonnormal components – see Phillips and Durlauf (1986), Park and Phillips (1988, 1989) and Sims et al. (1990). In this case the full matrix of estimated regression coefficients in a VAR is asymptotically normal but singular to the extent that there are some components in the system like unit roots and cointegrating vectors that converge at a faster rate. Since the estimated VAR coefficients are consistent (and, indeed, converge at faster rates in some directions) it might reasonably be expected that the impulse responses are also. However, as we show below, this is not the case for long horizon impulses. Moreover, since the estimated VAR coefficients have a limit normal distribution, albeit singular, it might also be anticipated that the impulse responses would be asymptotically normal. Again, we show this not to be the case. There are therefore some major differences in the limit theory of impulse responses between stationary and nonstationary VARs and these differences do seem to be important in the analysis of empirical results.

Let y_t be a m-vector time series generated by the following pth order VAR model

$$y_t = J(L)y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{1}$$

where $J(L) = \sum_{i=1}^{p} J_i L^{i-1}$. The system (1) is initialized at $t = -p+1, \ldots, 0$ and we may let these initial values be any random vectors including constants. It is often convenient to set the initial conditions so that the I(0) component of (1) is stationary and we will proceed as if this has been done. The presence of deterministic components in (1) does not affect our conclusions in any substantive way, so we will proceed as if they are absent just to keep the derivations as simple as possible. It is also convenient to write (1) in levels and differences format as

$$y_t = Ay_{t-1} + \Psi(L)\Delta y_{t-1} + \varepsilon_t, \quad A = J(1),$$
$$\Psi(L) = \sum_{i=1}^{p-1} \Psi_i L^{i-1}, \quad \Psi_i = -\sum_{h=i+1}^{p} J_h. \tag{2}$$

To fix ideas for our subsequent analysis it is helpful to be specific about the roots of (1), the dimension of the cointegrating space and the form of the cointegrating vectors. We therefore assume the following.

Assumption 2.1 (Reduced rank regression).

(a) ε_t is iid with zero mean, variance matrix $\Sigma_\varepsilon > 0$ and finite fourth cumulants.

(b) The determinantal equation $|I_m - J(L)L| = 0$ has roots on or outside the unit circle.

(c) $A = I + \alpha\beta'$ where α and β are $m \times r$ matrices of full column rank r. Without loss of generality, it will be assumed that β is orthonormal.

(d) $\alpha_\perp'(\Psi(1) - I_m)\beta_\perp$ is nonsingular, where $\alpha_\perp$ and $\beta_\perp$ are $m \times (m-r)$ matrices of full column rank that are orthogonal to α and β, respectively.

These conditions ensure that (2) has a reduced rank regression format and is the error correction model (ECM) of a system with some unit roots and some stationary components – see Toda and Phillips (1993) for further discussion. In place of condition (c) above we will also make use of the following weaker condition, which allows for some roots to be near unity.

(c′) $A = \beta_\perp \exp(n^{-1}\Gamma)\beta_\perp' + \beta\beta' + \alpha\beta'$ where α and β are $m \times r$ matrices of full column rank r, β is normalized to be orthonormal, and Γ is a constant matrix of dimension $s \times s$ with $s = m - r$.

We write the model (1) in companion form as

$$Y_t = CY_{t-1} + \Xi_t, \quad \Xi_t' = [\varepsilon_t', 0, \ldots, 0], \quad Y_t', = [y_t', \ldots, y_{t-p+1}'], \tag{3}$$

where

$$C = \begin{bmatrix} J_1 & \cdots & J_{p-1} & J_p \\ I & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & I & 0 \end{bmatrix}. \tag{4}$$

Let $M' = [I_m, 0, \ldots, 0]$, and then, up to initial conditions and deterministic components, the moving average (MA) representation of the system is

$$Y_t = \sum_{i=0}^{t-1} C^i \Xi_{t-i}, \quad \text{or } y_t = \sum_{i=0}^{t-1} M'C^iM\varepsilon_{t-i} = \sum_{i=0}^{t-1} \Theta_i \varepsilon_{t-i}, \quad \text{say.} \tag{5}$$

The system's impulse responses are given by the elements of the sequence of matrices Θ_i or certain linear combinations of the components of Θ_i, depending on the information that is supplied concerning the ordering of the shocks or structural relations among them. We will not be concerned with these latter issues as they do not affect the limit theory in any material way (unless, of course, there is lack of identification in relations among the shocks, in which case we would need to proceed as in the analysis of partially identified systems – see Phillips, 1989). Instead our primary interest is the behaviour of these impulse responses as the lead time $i \to \infty$, and the asymptotic behaviour of estimates of these quantities as $n \to \infty$.

Lemma 2.2. Under Assumption 2.1 *as* $i \to \infty$

$$\Theta_i \to \bar{\Theta} = \beta_\perp \beta_\perp' + \beta_\perp E_{12}(I - E_{22})^{-1} Q = \beta_\perp \beta_E', \quad say, \tag{6}$$

where $Q' = [\beta, H, 0, \ldots, 0]$ *is* $m \times (m(p-1)+r)$, *and where* E_{12} *and* E_{22} *are submatrices of the companion matrix* E *given in* (A.3) *below for the transformed system* (2′), *i.e. system* (2) *above rotated into separate* I(1) *and* I(0) *subsystems. For the case of near unit roots as in assumption* 2.1(c′), *when* $i = fn$ *where* $f > 0$ *is a fixed fraction of the sample we have as* $n \to \infty$

$$\Theta_i \to \bar{\Theta}_\Gamma = \beta_\perp \exp(f\Gamma)\beta_E'. \tag{6′}$$

According to (6) and (6′), the limiting impulse responses are nonzero only in those directions where the model is nonstationary and has unit roots or near unit roots, i.e. $\beta_\perp$. The limiting response matrices $\bar{\Theta}$ and $\bar{\Theta}_\Gamma$ both lie in the range of $\beta_\perp$. The fact that these matrices are nonzero has some important implications for inference, as our next result shows.

Theorem 2.3. Let Assumption 2.1 *hold and let* $\hat{\Theta}_i$ *be the OLS estimates of the impulse response matrices* Θ_i *in the MA representation* (5):

(i) *For fixed i we have*: $\hat{\Theta}_i \xrightarrow[p]{} \Theta_i$, $n^{1/2}(\hat{\Theta}_i - \Theta_i) \Rightarrow \mathrm{N}(0, V_i)$ *as* $n \to \infty$, *where*

$$V_i = N_i V_a N_i', \qquad N_i = \sum_{j=0}^{i-1} \Theta_{i-1-j} \otimes M' C'^j K^{-1},$$

$$V_a = \Sigma_\varepsilon \otimes G_\xi \Sigma_{\xi\xi}^{-1} G_\xi'.$$

Here, $\Sigma_{\xi\xi} = E(\xi_t \xi_t')$, *where* $\xi_t = [y_{t-1}'\beta, \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}']'$ *is the vector of stationary components in the system*; *and* G_ξ *is the matrix*

$$G_\xi = \begin{bmatrix} \beta & 0 \\ 0 & I_{m(p-1)} \end{bmatrix},$$

and K is the matrix that transforms (1) *into* (2). *Section* A.2 *below defines these matrices and details these transformations. The symbol '$\Rightarrow$' signifies weak convergence and we use the convention that the matrix normal distribution is written in terms of stacked rows of the matrix variate.*

(ii) *If* $i/n \to 0$ *as* $n, i \to \infty$, *we have*: $\hat{\Theta}_i \underset{p}{\longrightarrow} \bar{\Theta}$ *as* $n \to \infty$.

(iii) *If* $i = fn$ *where* $f > 0$ *is a fixed fraction of the sample, we have*

$$\hat{\Theta}_i \Rightarrow \beta_\perp \exp(fU)\beta_\perp' + \beta_\perp \exp(fU)E_{12}(I - E_{22})^{-1}Q$$

$$= \beta_\perp \exp(fU)\beta_{E'}' \quad say,$$

where U is random and has a matrix unit root distribution (given explicitly in (A.6) *below*).

(iv) *If condition* (c$'$) *replaces condition* (c) *and* $i = fn$ *with* $f > 0$ *fixed, then*

$$\hat{\Theta}_i \Rightarrow \beta_\perp \exp(fU_\Gamma)\beta_\perp' + \beta_\perp \exp(fU_\Gamma)E_{12}(I - E_{22})^{-1}Q$$

$$= \beta_\perp \exp(fU_\Gamma)\beta_E',$$

where U_Γ *is random and has a matrix local-to-unity distribution (see* (A.6) *below*).

Remark 2.4. Theorem 2.3 shows that when there are unit roots or near unit roots in a VAR system, the long period ahead impulse responses estimated by an unrestricted OLS regression are inconsistent. In particular, the limits of the estimated responses are random variables rather than the true impulse responses. This may seem surprising given that the presence of unit roots or near unit roots accelerates the convergence of the coefficient estimates in an OLS regression – on that basis one might have anticipated that the impulse responses would, if anything, converge faster in some directions. The reason for the inconsistency is that the true impulse responses no longer die out as the lead time increases, i.e. the elements of Θ_i do not tend to zero as $i \to \infty$, but carry the effects of the unit roots with them indefinitely. However, the unit roots are estimated with error and the effects of the estimation error persist in the limit as $n \to \infty$ when we consider long period ahead impulses Θ_i where i is some fraction (f) of the sample size (n). By contrast, when the system is stable the elements of Θ_i tend to zero as $i \to \infty$, and the estimation errors have no effect in the limit. In this case, $\beta_\perp$ is null, $\bar{\Theta} = 0$, and $\hat{\Theta}_i \underset{p}{\longrightarrow} 0$. Thus, (ii)–(iv) cover the stable case as well.

Remark 2.5. For fixed i we get asymptotic normal distributions for the impulse response matrices $\hat{\Theta}_i$, just as we do in the stationary case – see Lütkepohl (1993) for the latter. However, the limit variance matrix V_a of the estimated VAR coefficients that enter the formula is now singular because only the stationary components of the system contribute to these $\sqrt{n}$-asymptotics.

Example 2.6. To illustrate the formulae, take the special case of the random-walk model $y_t = ay_{t-1} + \varepsilon_i$, $a = 1$. i.e. set $m = 1$ and $p = 1$ in (1). Then the estimated impulse responses are $\hat{a}^i$ and when $i = fn$ we have

$$\hat{a}^i = \left(1 + \frac{n(\hat{a}-1)}{n}\right)^i$$
$$\Rightarrow \exp\left\{ f\left(\int_0^1 S\,\mathrm{d}S\right)\left(\int_0^1 S^2\right)^{-1}\right\} \quad \text{as } n \to \infty, \text{ where } S \equiv BM(1).$$

Here, the lead time is a fraction of the sample size, and the estimated impulse responses effectively exponentiate a random variable in the vicinity of unity and therefore tend to an exponential unit-root distribution. When the lead time i is fixed, we get $\hat{a}^i \xrightarrow[\text{p}]{} 1$ instead, but then

$$n(\hat{a}^i - 1) = in(\hat{a} - 1) + O_p(n^{-1}) \Rightarrow i\left(\int_0^1 S\,\mathrm{d}S\right)\left(\int_0^1 S^2\right)^{-1}.$$

and the limit distribution is proportional to a unit-root distribution and is again asymmetric. In this case, G_ξ is a null matrix (there are no stationary components in the system) and $V_a = 0$ in part (i) of the theorem. In both cases, the limit theory is nonnormal.

Fig. 1(a) shows the limit distributions of $\hat{a}^i$ when $i = fn$ for various values of f. The asymmetry of the distributions of the estimated impulse responses is similar to that of the usual unit-root distribution, but the support of the distribution is the positive half-line rather than the whole real axis. The asymmetry is strongly evident when $f = 0.25$ and less marked when $f = 0.02$. The asymmetry of estimated impulse response distributions has been noted in some simulation work previously, and is often attributed to the nonlinearity of the impulse responses. This has led to some research on ways to adjust confidence regions for the impulse responses to take account of the asymmetry (e.g. Quah and Blanchard, 1993; Sims, 1994). The above limit theory shows that in cases where there are unit roots or near unit roots, the reason for the asymmetry in the distribution is, in fact, the nonnormal asymmetric limit theory of the estimated impulse responses. As Theorem 2.3 shows, these nonnormal asymptotics dominate the distributional shape of the estimated impulse responses even when there are stationary components in the system.

Example 2.7. Next consider the AR(2) model $y_t = ay_{t-1} + b\Delta y_{t-1} + \varepsilon_t$, $a = 1$, $|b| < 1$. The impulse responses of this system are $\theta_i = 1 + b + \cdots + b^i \to (1-b)^{-1}$. The estimated responses are

$$\hat{\theta}_i = \hat{a}^i + \hat{a}^{i-1}\hat{b} + \cdots + \hat{b}^i$$

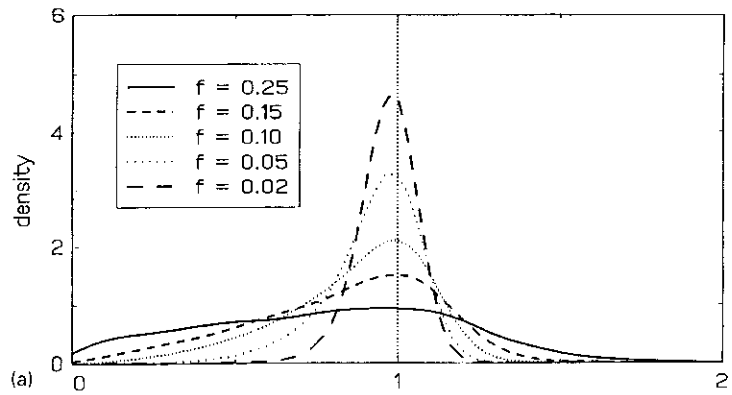

Fig. 1(a). Limit density of OLS impulse response h-periods ahead for a random walk: $h = f^* n$.

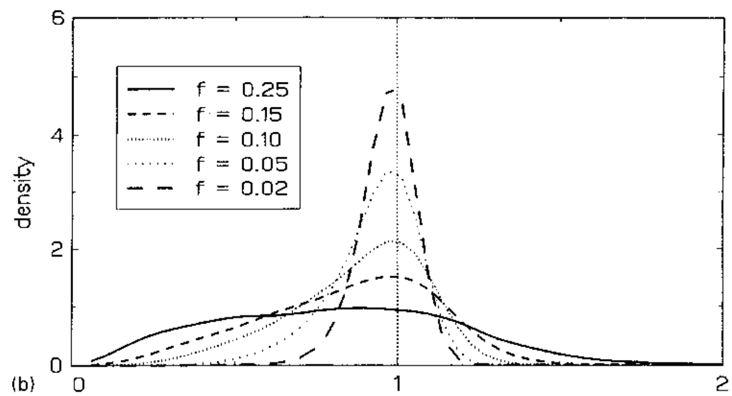

Fig. 1(b). Limit density of OLS forecast error variance h-periods ahead for a random walk: $h = f^* n$.

$$\left\{ \begin{array}{ll} \xrightarrow[p]{} \theta_i, & \text{for } i \text{ fixed} \\ \Rightarrow \exp\left\{ f\left(\int_0^1 S\,\mathrm{d}S\right)\left(\int_0^1 S^2\right)^{-1}\right\}(1-b)^{-1} & \text{for } i = fn \end{array} \right\} \text{ as } n \to \infty.$$

Again, the impulse responses are inconsistent and have random limits. Note that the stationary coefficient (b) in this system does figure in the limit distribution. Because of the unit root, all of the stationary components are accumulated and this leads to the presence of the scaling factor $1/(1-b)$ in the above limit.

Reduced rank regression impulse responses 2.8. If (2) is estimated as a VAR of reduced rank then the limit theory for the impulse responses is different. We may assume that the rank of the system is either known a priori to be r or is consistently estimated (e.g. by the order selection technique in Phillips, 1994; Chao and Phillips, 1994). The case where the rank of the system is incorrectly specified can also be analysed, of course, and the unrestricted VAR regression considered above is one instance of this. However, as we have seen, in this case the long period ahead responses are inconsistently estimated. In the case where the rank is correctly specified (or consistently estimated), the fact that the system has $s=m-r$ unit roots is also known (or consistently estimated) and this knowledge then becomes part of the model. It turns out to have a pivotal influence on the asymptotic theory. In a reduced rank regression the matrix product $\alpha\beta'$ is estimated in place of an unrestricted coefficient matrix for the lagged levels variable in (2). In consequence, no unit roots are estimated (either explicitly or implicitly), and this affects the limit theory for the system's estimated impulse responses in a material way.

Theorem 2.9. Let Assumption 2.1 *hold and let* $\hat{\Theta}_i$ *be the estimates of the impulse response matrices* Θ_i *obtained from a reduced rank regression on* (2).

(i) *For fixed i we have:* $\hat{\Theta}_i \underset{p}{\longrightarrow} \Theta_i$, *and* $n^{1/2}(\hat{\Theta}_i-\Theta_i)\Rightarrow \mathrm{N}(0,V_i)$ *as* $n\to\infty$, *where*

$$V_i=N_iV_aN_i', \qquad N_i=\sum_{j=0}^{i-1}\Theta_{i-1-j}\otimes M'C'^{j}K^{-1},$$

$$V_a=\Sigma_{\varepsilon\varepsilon}\otimes G_\xi\,\Sigma_{\xi\xi}^{-1}\,G_\xi',$$

where $\Sigma_{\xi\xi}=E(\xi_t\xi_t')$, *and* $\xi_t=[y_{t-1}'\beta,\Delta y_{t-1}',\dots,\Delta y_{t-p+1}']'$ *and* G_ξ *are as in Theorem* 2.3.

(ii) *If* $i\to\infty$ *as* $n\to\infty$ *with either* $i=fn$ *or* $i/n\to 0$, *we have:* $\hat{\Theta}_i \underset{p}{\longrightarrow} \bar{\Theta}$ *and* $n^{1/2}(\hat{\Theta}-\bar{\Theta})\Rightarrow \mathrm{N}(0,\bar{V})$ *as* $n\to\infty$, *where*

$$\bar{V}=NV_aN', \qquad N=\bar{\Theta}\otimes[\beta,H,0,\dots,0](I-E_{22}')^{-1}\begin{bmatrix} I_r & 0 \\ 0 & I_{p-1}\otimes H' \end{bmatrix}.$$

Remark 2.10. Theorem 2.9(i) shows that the estimated impulse responses in a cointegrated VAR model are consistent when they are based on a reduced rank regression in which the cointegrating rank is consistently estimated. The result shows that it is important in a reduced rank regression to estimate the cointegrating rank by a consistent method. Order selection methods like those used in Phillips (1994) are one possibility here. Another is to use classical likelihood ratio tests, as in Johansen (1988, 1991), that are suitably modified to ensure that the size of the test goes to zero as the sample size goes to infinity. The consistency of the estimated impulse responses applies to both short period and long period ahead responses. In the latter case, shocks have a persistent effect on the system indefinitely into the future. It is these persistent effects that the reduced rank regression estimates consistently. Part (ii) of the theorem shows that when there are near unit roots in the system, the reduced rank regression mistakenly takes roots near unity as roots at unity when $n \to \infty$. The same is true when model selection criteria such as BIC and PIC are used to select cointegrating rank. In consequence, the reduced rank regression long period impulse responses are inconsistently estimated when there are near unit roots. The limit distribution of these estimated impulses is still Gaussian.

Example 2.11. As an illustration, consider the model

$$\begin{aligned} y_{1t} &= y_{1t-1} + \varepsilon_{1t}, \\ y_{2t} &= b y_{1t-1} + \varepsilon_{2t}, \end{aligned} \quad \text{i.e. } y_t = A y_{t-1} + \varepsilon_t,$$

with

$$A = \begin{bmatrix} 1 & 0 \\ b & 0 \end{bmatrix} = I_2 + \alpha\beta'; \quad \alpha = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} b \\ -1 \end{bmatrix}.$$

The impulse responses are $\Theta_i = A^i = A$, $\forall i$. The estimated coefficient matrix from a reduced rank regression is $\hat{A} = I + \hat{\alpha}\hat{\beta}'$, and the associated impulse responses are $\hat{A}^i$. Let $\hat{H} = [\hat{\beta}_\perp^n \ \hat{\beta}^n]$ where the superscript n signifies the normalized vector. Then

$$\hat{H}'\hat{A}\hat{H} = \begin{bmatrix} 1 & \hat{\beta}_\perp'\hat{\alpha} \\ 0 & 1 + \hat{\beta}'\hat{\alpha} \end{bmatrix} \quad \text{and as } n \to \infty$$

(with $i \to \infty$ such that $in^{-1} \to 0$, or $i = fn$)

$$\hat{H}'\hat{A}^i\hat{H} = \begin{bmatrix} 1 & \hat{\beta}_\perp'\hat{\alpha}\sum_{j=0}^{i-1}(1 + \hat{\beta}'\hat{\alpha})^j \\ 0 & (1 + \hat{\beta}'\hat{\alpha})^i \end{bmatrix} \underset{p}{\longrightarrow} \begin{bmatrix} 1 & \beta_\perp'\alpha\{1 - (1 + \beta'\alpha)\}^{-1} \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & b \\ 0 & 0 \end{bmatrix}.$$

Thus,

$$\hat{A}^i \underset{p}{\longrightarrow} H \begin{bmatrix} 1 & b \\ 0 & 0 \end{bmatrix} H' = (1 + b^2)^{-1}\{\beta_\perp\beta_\perp' + \beta_\perp b\beta'\}$$

$$= (1 + b^2)^{-1}\left\{\begin{bmatrix} 1 & b \\ b & b^2 \end{bmatrix} + \begin{bmatrix} b^2 & -b \\ b^3 & -b^2 \end{bmatrix}\right\} = \begin{bmatrix} 1 & 0 \\ b & 0 \end{bmatrix},$$

giving consistent estimates of the impulse responses in the limit. The limit of the OLS estimated impulse response matrix on the other hand is the random matrix

$$e^{fU}\begin{bmatrix} 1 & 0 \\ b & 0 \end{bmatrix}.$$

Remark 2.12. In the above example, the cointegrating coefficient b could be estimated directly by regression methods, such as fully modified least squares – see Phillips and Hansen (1990). The limit theory for the estimated impulse responses is the same in this case as it is for reduced rank regression. More generally, models such as that of Example 2.11 are explicit error correction models (ECMs), where the number and location of the unit roots and cointegrating vectors is given as part of the specification of the system. In such cases we get the same asymptotic theory for ECM estimators like FM-OLS as that given in Theorem 2.9 for the reduced rank regression. Thus, all these procedures share the same advantage over unrestricted VAR regression that they deliver consistent estimates of the impulse responses.

Remark 2.13. The case of a reduced rank regression in the presence of near unit roots can be handled in the same way as Theorem 2.9. In such cases, model selection procedures like BIC and PIC will mistakenly take roots that are near unity as roots at unity, at least in large samples. In this event, although the true impulse response matrices behave as in (6′), the reduced rank regression estimates will satisfy $\hat{\Theta}_i \underset{p}{\longrightarrow} \bar{\Theta}$, as in Theorem 2.9.

3. Forecast error variance asymptotics

The limit theory for the estimated impulse response matrices can be used to deliver forecast error variance asymptotics. From (5), the forecast error of the optimal h-step ahead predictor, $y_{t,h}$, and its variance matrix are:

$$y_{t+h} - y_{t,h} = \sum_{i=0}^{h-1} \Theta_i \varepsilon_{t+h-i},$$

$$\mathrm{FEV}(y_{t,h}) = E(y_{t+h} - y_{t,h})(y_{t+h} - y_{t,h})' = \sum_{i=0}^{h-1} \Theta_i \Sigma_\varepsilon \Theta_i'$$

$$= \sum_{i=0}^{h-1} \Phi_i \Phi_i' = F(h) \quad \text{say},$$

where $\Phi_i = \Theta_i P$ and P is a lower triangular matrix from the Choleski decomposition of $\Sigma = PP'$. By virtue of Lemma 2.2 and Cesaro summation we have

$$h^{-1}F(h) \to \bar{\Theta}\Sigma_\varepsilon\bar{\Theta}' = \bar{F} \quad \text{say, and therefore } F(h) \sim h\bar{\Theta}\Sigma_\varepsilon\bar{\Theta}' \quad \text{as } h \to \infty.$$

Let $\varphi_{i,jk}$ be the jkth element of Φ_i. Then

$$\tau_{jk,h} = h^{-1} \sum_{i=0}^{h-1} \varphi_{i,jk}^2$$

is the contribution to the h-step forecast error variance of the jth variable in the system that is due to the (orthonormalized) innovations in variable k. Similarly,

$$\omega_{jk,h} = \tau_{jk,h} \Big/ \sum_{k=1}^{m} \tau_{jk,h}$$

is the proportion of the overall forecast error variance in variable j that is due to variance k. These quantities are the critical elements in the forecast error variance decomposition of a VAR model. They are used extensively in empirical work for policy analysis purposes to determine the effects of unanticipated shocks to one variable on other variables in the system over time. The following result gives the limit theory for estimates of these quantities obtained from an unrestricted VAR regression.

Theorem 3.1. Let Assumption 2.1 *hold and let* $\hat{\Theta}_i$ *be the OLS estimates of the impulse response matrices* Θ_i *and* $\hat{\Sigma}_\varepsilon$ *be the OLS estimate of the equation error variance matrix. Denote by* $\hat{F}(h)$ *the corresponding estimate of the forecast error variance matrix* $F(h)$.
(i) *For fixed h we have*: $\hat{F}(h) \xrightarrow[p]{} F(h)$, $n^{1/2}(\hat{F}(h) - F(h)) \Rightarrow \mathrm{N}(0, V_h)$ *as* $n \to \infty$, *where*

$$V_h = N_{ah} V_a N_{ah}' + N_\sigma V_\sigma N_\sigma',$$

$$N_{ah} = \sum_{i=0}^{h-1} [(I \otimes \Theta_i \Sigma_\varepsilon) + (\Theta_i \Sigma_\varepsilon \otimes I) K_{mm}] N_i$$

$$N_i = \sum_{j=0}^{i-1} \Theta_{i-1-j} \otimes M' C'^j K^{-1},$$

$$N_\sigma = \sum_{i=0}^{h-1} (\Theta_i \otimes \Theta_i) D,$$

$$V_a = \Sigma_\varepsilon \otimes G_\xi \Sigma_{\xi\xi}^{-1} G_\xi' \quad \textit{and} \quad V_\sigma = D^+ (\mathrm{var}(\varepsilon_t \otimes \varepsilon_t)) D^{+\prime}.$$

The above formulae employ the following notation: D *is the duplication matrix for which* $\mathrm{vec}(A) = Da$, *where* a *is the vector of nonredundant elements of a symmetric matrix* A; *the matrix* $D^+ = (D'D)^{-1} D'$ *is a generalized inverse of* D; K_{mm} *is the commutator matrix for which* $K_{mm} \mathrm{vec}(X) = \mathrm{vec}(X')$, *where* X *is an arbitrary* $m \times m$ *matrix. The other notation is the same as that defined in Theorem* 2.3. *If the errors in* (2) *are normally distributed, then*

$\text{var}(\varepsilon_t \otimes \varepsilon_t) = 2P_D(\Sigma_\varepsilon \otimes \Sigma_\varepsilon)$, *where* $P_D = D(D'D)^{-1}D'$, *and the covariance matrix* V_σ *is simply* $2D^+(\Sigma_\varepsilon \otimes \Sigma_\varepsilon)D^{+\prime}$.

(ii) *If* $h = fn$ *where* $f > 0$ *is a fixed fraction of the sample, we have*:

(a) $h^{-1}\hat{F}(h) \Rightarrow f^{-1}\int_0^f \beta_\perp e^{sU}\beta_E'\Sigma_\varepsilon\beta_E e^{sU}\beta_\perp' = V_F(U)$, *say, where* U *is the unit-root matrix variate given in* (A.6).

(b) $\tau_{jk,h} \Rightarrow f^{-1}\int_0^f (\beta_{j\perp}e^{sU}\beta_E'P_k)^2\,ds$,

(c) $\omega_{jk,h} \Rightarrow \int_0^f (\beta_{j\perp}e^{sU}\beta_E'P_k)^2\,ds / \int_0^f (\beta_{j\perp}e^{sU}\beta_E'\Sigma_\varepsilon\beta_E e^{sU'}\beta_{j\perp}'\,ds$, *where* $\beta_{j\perp}$ *is the* j*th row of* $\beta_\perp$ *and* P_k *is the* k*th column of* P.

(iii) *If condition* (c′) *replaces condition* (c) *in Assumption* 2.1 *and* $i = fn$ *with* $f > 0$ *fixed, then*

$$h^{-1}\hat{F}(h) \Rightarrow f^{-1}\int_0^f \beta_\perp e^{sU_\Gamma}\beta_E'\Sigma_\varepsilon\beta_E e^{sU_\Gamma'}\beta_\perp' = V_F(U_\Gamma),$$

where U_Γ *is the local-to-unity matrix variate given in* (A.6). *Similar changes occur in the limits given in* (ii) (b) and (c) *above.*

Remark 3.2. In models with nonstationary elements, we expect the forecast error variance to grow in a linear way with the forecast horizon. This is precisely what happens with the forecast error variance matrix of the optimal predictor. As shown above, the matrix $F(h) \sim h\bar{\Theta}\Sigma_\varepsilon\bar{\Theta}'$ as $h \to \infty$. In contrast, the estimated forecast error variance matrix from an estimated unrestricted VAR with some roots at or near unity behaves like a random matrix multiple of the lead time h rather than a constant matrix multiple of h, as shown in part (ii) (a) of the theorem. The expression for $V_F(U)$ shows that the random matrix is a continuous average of a matrix quadratic form in the limiting impulse responses. Thus, estimated forecast error variance matrices for long lead times in unrestricted VARs are inconsistent. The same conclusion follows for the corresponding estimates of the forecast error variance decompositions.

Example 3.3 (*Example* 2.6 *continued*). This is the scalar random-walk case, and when $h = fn$ we have

$$h^{-1}\hat{F}(h) \Rightarrow \sigma_\varepsilon^2 f^{-1}\int_0^f e^{2pu}dp = \sigma_\varepsilon^2\,\frac{e^{2fu} - 1}{2uf} = v_F \quad \text{say},$$

where $u = (\int_0^1 S\,dS)(\int_0^1 S^2)^{-1}$ is the scalar unit-root distribution. Note that when $f = 1$, we have the elementary inequality $e^x > 1 + x$, $\forall x \neq 0$ so that

$$P(v_F < \sigma_\varepsilon^2) = P(e^{2fu} - 1 > 2fu; u < 0) = P(u < 0) = P(\chi_1^2 < 1) = 0.68.$$

Since $\lim_{h\to 0} h^{-1}F(h) = \sigma_\varepsilon^2$, it follows that v_F underestimates the actual forecast error variance of the optimal predictor with a probability of 0.68 in the limit as $h \to \infty$. This means that unrestricted OLS regression estimates not only give inconsistent estimates of the forecast error variance of a random walk at long horizons, but also have a clear tendency to underestimate this variance.

Fig. 1(b) shows the limit distribution, v_F, of the forecast error variance when $h = fn$ for various values of f. The distributions are similar to those of the impulse responses. Again the asymmetry is strongest when f is largest.

Theorem 3.4. Let $\hat{y}_{n,h}$ be the h-step ahead forecast of y_{n+h} from an unrestricted VAR regression (1), *using sample data $t \leqslant n$. Under Assumption* 2 *and when $h = fn$ we have the following limit theory as $n \to \infty$:*

(i) $n^{-1/2} y_{n,h} \Rightarrow \beta_{\perp} \beta_E' S(1)$;
(ii) $n^{-1/2} \hat{y}_{n,h} \Rightarrow \beta_{\perp} \exp(fU) \beta_E' S(1)$;
(iii) $n^{-1/2}(y_{n+h} - \hat{y}_{n,h}) \Rightarrow \beta_{\perp}[I - \exp(fU)]\beta_E' S(1) + \beta_{\perp} \beta_E' S_+(f)$;
(iv) $n^{-1/2}(y_{n+h} - y_{n,h}) \Rightarrow \beta_{\perp} \beta_E' S_+(f)$.

In the above formulae, S and S_+ are independent Brownian motions with the same variance matrix given by $lr\,\mathrm{var}(u_{1t})$.

(v) *When* (c′) *replaces* (c) *in Assumption* 2, *the matrix U in the above limits is replaced with U_Γ, the unit-root matrix I is replaced with* $\exp(f\Gamma)$, *and the Brownian motions S and S_+ are replaced with* $J_\Gamma(r) = \int_0^r \exp((r-s)\Gamma)\,\mathrm{d}S$, *and* $J_\Gamma^+(f) = \int_0^r \exp((r-s)\Gamma)\,\mathrm{d}S_+$.

Remark 3.5. In a stationary VAR the forecast error of the optimal predictor is a random sequence that converges to a limit random vector as the forecast horizon tends to infinity. When there are nonstationary components in a VAR, the error in the optimal predictor behaves like a random walk, is of the same order as the square root of the forecast horizon, and when appropriately standardized it tends to a Brownian motion process, as shown in part (iv) of Theorem 3.4. Part (ii) of the theorem shows that the feasible predictor obtained from an estimated VAR does not have the same limit behaviour as that of the optimal predictor, but carries with it the effects of the estimated unit roots (or near unit roots) in the model. In consequence, the error in the feasible predictor has two independent components in the limit: one component is the same as that of the error in the optimal predictor; the other component measures the difference in the limit between the feasible and the optimal predictor and results from the estimated nonstationary components in the model. Hence, the latter component figures only in nonstationary directions, as is apparent from the form of the limit shown in part (iii) of the theorem. The upshot of this result is that prediction from an unrestricted nonstationary VAR regression is not asymptotically optimal in the sense that the predictions do not converge to the optimal predictors, at least over long forecast horizons. This result is in direct contrast to that of a stationary VAR, where the coefficient estimation errors have no effects asymptotically and the difference between the feasible and optimal predictors tends to zero as $n \to \infty$.

Reduced rank regression asymptotics 3.6. As shown in Sections 2.8–2.9, the limit theory for impulse responses that are estimated by a reduced rank regression differs from that of an unrestricted VAR. As a consequence, forecast error variance asymptotics also differ. When the cointegrating rank of the VAR is consistently estimated, then so are the forecast error variances and forecast error variance decompositions.

Theorem 3.7. Let Assumption 2.1 *hold and assume that* (2) *is estimated by a reduced rank regression with cointegrating rank r either known a priori or consistently estimated as in Chao and Phillips* (1994). *Let* $\hat{\Theta}_i$ *be the estimates of the impulse response matrices* Θ_i, *and* $\hat{\Sigma}_\varepsilon$ *be the estimate of the equation error variance matrix obtained from the residuals of the reduced rank regression. Denote by* $\hat{F}(h)$ *the corresponding estimate of the forecast error variance matrix* $F(h)$.

(i) *For fixed h we have*: $\hat{F}(h) \xrightarrow[p]{} F(h)$, $n^{1/2}(\hat{F}(h) - F(h)) \Rightarrow \mathrm{N}(0, V_h)$ *as* $n \to \infty$, *where the variance matrix* V_a *is the same as that given in Theorem* 3.1 (i).
(ii) *If* $h \to \infty$ *as* $n \to \infty$ *with either* $h = fn$ *or* $h/n \to 0$, *we have*

$$h^{-1}\hat{F}(h) \xrightarrow[p]{} \overline{F}, \qquad \hat{\tau}_{jk,h} \xrightarrow[p]{} \overline{\tau}_{jk} \quad \text{and} \quad \hat{\omega}_{jk,h} \xrightarrow[p]{} \overline{\omega}_{jk}.$$

Here $\overline{\tau}_{jk} = \lim_{h\to\infty} h^{-1}\sum_{i=0}^{h-1} \varphi_{i,jk}^2 = \overline{\varphi}_{jk}^2$, *and* $\overline{\omega}_{jk} = \overline{\tau}_{jk}/\sum_{k=1}^{m} \overline{\tau}_{jk}$, *where* $\overline{\varphi}_{jk}$ *is the jkth element of* $\overline{\Phi} = \bar{\Theta}P$.

Remark 3.8. Theorem 3.7 shows that fixed period horizon decompositions of the forecast error variances that are estimated from a reduced rank regression have the same limit theory as unrestricted VAR estimates. But when the forecast horizon tends to infinity with the sample size, the reduced rank regression estimates continue to be consistent, whereas those from an unrestricted VAR are inconsistent and have random limits.

Theorem 3.9. Let $\hat{y}_{n,h}$ *be the h-step ahead forecast of* y_{n+h} *from a reduced rank regression on* (2), *using sample data* $t \leqslant n$. *Under Assumption* 2 *and when* $h = fn$ *we have the following limit theory as* $n \to \infty$:

(i) $n^{-1/2}y_{n,h}, n^{-1/2}\hat{y}_{n,h} \Rightarrow \beta_\perp \beta_E' S(1)$;
(ii) $n^{-1/2}(y_{n+h} - y_{n,h}), n^{-1/2}(y_{n+h} - \hat{y}_{n,h}) \Rightarrow \beta_\perp \beta_E' S_+(f)$.

When (c′) *replaces* (c) *in Assumption* 2 *we have*

(iii) $n^{-1/2}\hat{y}_{n,h} \Rightarrow \beta_\perp \beta_E' J_\Gamma(1)$,
(iv) $n^{-1/2}(y_{n+h} - y_{n,h}) \Rightarrow \beta_\perp \beta_E' J_\Gamma^+(f)$,
(v) $n^{-1/2}(y_{n+h} - \hat{y}_{n,h}) \Rightarrow \beta_\perp[\exp(f\Gamma) - I]\beta_E' J_\Gamma(1) + \beta_\perp \beta_E' J_\Gamma^+(f)$.

The notation in the above formulae is defined in Theorem 3.4.

Remark 3.10. Parts (i) and (ii) of Theorem 3.9 show that forecasts from a correctly specified reduced rank regression are asymptotically equivalent to the optimal predictor. Again, the same result holds if a consistent estimate of the cointegrating rank is used in the fitted model. However, when the system has near unit roots instead of unit roots, the reduced rank regression forecasts are no longer asymptotically equivalent to those of the optimal predictor. As is apparent from the first term on the right-hand side of part (v), within sample estimation errors figure in the limit formula for the error in the reduced rank regression forecasts. Thus, we lose the consistency of the reduced rank regression predictor for the optimal predictor as the model itself drifts because of the presence of near unit roots and the localizing parameter matrix Γ is not consistently estimable.

4. Simulation evidence

4.1. DGP, impulse responses and FEVD

Some small-scale simulations were conducted to assess the accuracy of forecasting and policy analysis of models in the VAR class. We used the following data generating mechanism so that we could focus attention on the effects of a unit root and cointegration on the forecasts and impulse responses.

$$y_t = Ay_{t-1} + \varepsilon_t, \qquad \varepsilon_t = \mathrm{iidN}(0, I_3),$$
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} = I_3 + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix}. \tag{7}$$

This system has one unit root (in the first equation) and two cointegrating vectors. The impulse responses are:

$$A^{i} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \forall i \geqslant 2; \tag{8}$$

and the forecast error variance matrix is

$$F(h) = \sum_{i=0}^{h-1} A^{i}A'^{i} = h \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} [\,1 \quad 2 \quad -1\,]$$
$$\text{so that } h^{-1}F(h) = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ -1 & -2 & 1 \end{bmatrix}.$$

The VAR estimated impulse responses, and estimated FEV and the FEVD quantities have the following limits:

$$\hat{\Theta}_i \Rightarrow \mathrm{e}^{fU}\beta_\perp\beta_E' = \mathrm{e}^{fU}\begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix},$$

$$h^{-1}\hat{F}(h) \Rightarrow \left(f^{-1}\int_0^f \mathrm{e}^{2sU}\,\mathrm{d}s\right)\beta_\perp\beta_E'\Sigma_\varepsilon\beta_E\beta_\perp'$$

$$= \sigma_{11}\frac{\mathrm{e}^{2fu}-1}{2u}\beta_\perp\beta_\perp' = v_F\begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ -1 & -2 & 1 \end{bmatrix},$$

$$\hat{\tau}_{jk,h} \Rightarrow v_F\beta_{j\perp}\delta_{lk} = \begin{cases} v_F\beta_{j\perp}, & k=1, \\ 0, & k\neq 1, \end{cases}$$

$$\hat{\omega}_{jk,h} \Rightarrow (\beta_{j\perp}\delta_{lk})^2/(\beta_{j\perp})^2 = \begin{cases} 1, & k=1, \\ 0, & k\neq 1. \end{cases} \tag{9}$$

Note that the VAR estimated FEVD quantity $\hat{\varpi}_{jk,h}$ has a nonrandom limit in this case, and is consistent. Also, 100% of the FEVD for each of the 3 variables in the model is due to error 1 in the limit, and this is explained by the fact that the error in the first equation is the only persistent error in the model. Finally, note that the VAR estimated FEVD is consistent in this case – the random component that arises from the estimated unit root in the model scales out of the numerator and denominator of the FEVD formula.

4.2. *Model classes, model selection and parameter settings*

Our simulation experiments employed the range of models listed below for comparative purposes. In the Bayesian vector autoregression (BVAR) models we used a pre-set trend degree $t=0$ (i.e. an intercept was included in the regression), a uniform prior on the intercept, and a Minnesota prior (see Litterman, 1986; Todd, 1990) on the AR coefficients with both Litterman (designated as 'lit') and data-determined (designated as 'opt') settings for the tightness hyperparameter. The data-determined hyperparameters were selected using the predictive PIC criterion given in Phillips (1994, Eq. (45)) applied to the hyperparameters over the following intervals: $\lambda \in [0.01, 0.60]$ for the general tightness hyperparameter; and $\theta \in [0.01, 1.00]$ for the cross variable hyperparameter in the symmetric Minnesota prior. Lag length, trend degree and cointegrating rank were all data-determined using the predictive PIC criterion in the reduced rank regression (RRR) model and the error correction model (ECM) – again see Phillips (1994) for details of the implementation of this model determination criterion. The models are:

1. **VAR(p) + Tr(t):** A VAR model with trend degree t and lag length p, both determined by predictive PIC.
2. **BVAR(lit & opt):** BVAR models with pre-set trend degree $t=0$, uniform prior on the intercept, and a symmetric Minnesota prior on the matrices of AR coefficients using both Litterman(lit) and data-determined(opt) settings for the tightness hyperparameters.
3. **RRR:** a VAR(p) + Tr(t) model with lag-one coefficient matrix of possible reduced rank(r) to allow for cointegration among the variables. Lag length(p), trend degree(t) and cointegrating rank(r) are all data-determined by predictive PIC.
4. **ECM:** a VAR(p) + TR(t) model formulated in differences with a coefficient matrix on the lag-one levels variable that allows for cointegration of the specific form given in the DGP above (i.e., the structural component of the model is assumed to be correctly specified). The lag length (p) and the trend degree (t) are determined by using predictive PIC.

Our settings for the maximum lag length and trend degrees in model classes 1, 4 and 5 above are as follows: lag length, $p\max = 4$; trend degree, $t\max = 1$. In the BVAR models we set the parameters to $p=4$, $t=0$. Past experience with BVAR models in forecasting has shown that the inclusion of a linear trend generally causes a deterioration in forecasting performance – some recent evidence is reported in Phillips (1992, 1995a). Our setting of $t=0$ in the BVAR models reflects this experience and is designed to make the BVAR results more realistic from this perspective. In the other models, the trend degree is selected using predictive PIC, and we therefore allowed for a search over the cases of no intercept ($t=-1$), intercept ($t=0$) and linear trend ($t=1$).

We ran 1000 replications of 112 sample observations generated by the system (6). For each replication the parameters of the various models described in the preceding section were estimated from the first 100 sample observations, forecasts were generated up to 12 periods ahead, and impulse response coefficients (up to 30 periods) and forecast error decompositions were calculated.

4.3. *Forecasting results*

The forecasting results are shown in Figs. 2(a)–(c). These figures plot the average forecast root mean squared errors (RMSEs) over the 1000 replications for forecasts of the three variables obtained by the methods described above. Specifically, we calculated forecasts and performed policy analyses using: (i) unrestricted vector autoregression (denoted OLS); (ii) restricted ECM estimation (denoted ECM); (iii) reduced rank regression (denoted RRR); (iv) and (v) Bayesian vector autoregression with Litterman settings for the Minnesota

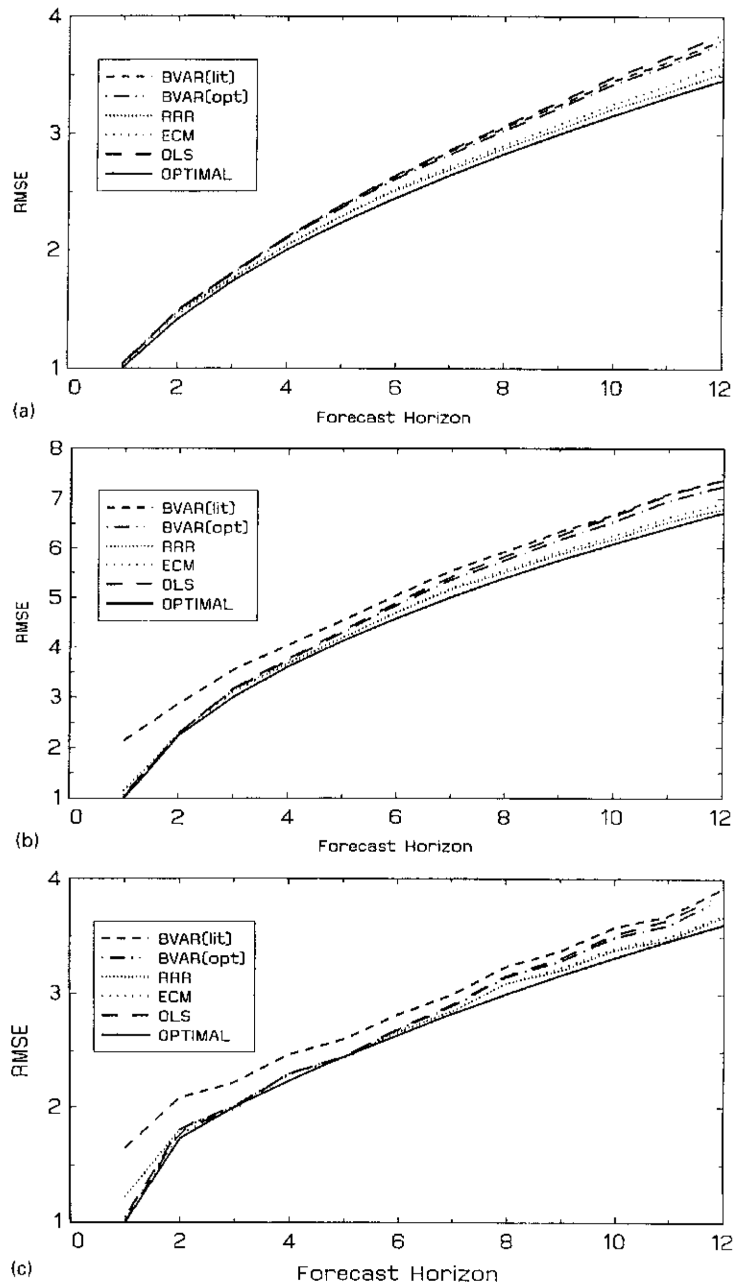

Fig. 2. (a) Average forecast RMSEs: equation: $Y_{1t}=Y_{1t-1}+e_{1t}$. (b) Average forecast RMSEs: Eq.: $Y_{2t}=2Y_{1t-1}+e_{2t}$. (c) Average forecast RMSEs: Eq.: $Y_{3t}=Y_{1t-1}-Y_{2t-1}+e_t$.

priors (denoted BVAR(lit)), and optimum data-determined settings for the hyperparameters obtained by the predictive PIC criterion (denoted BVAR(opt)). The graphs show the forecast RMSEs from the simulations with these estimated models against those of the optimal predictor (which is calculated analytically for the above system (6) with the true parameter settings). The latter graph (shown as the solid line in the figures) represents the optimal forecast envelope for this system. The conclusions to emerge are as follows:

(i) The data-determined RRR and ECM models produce better forecasts on average than the BVARs and the unrestricted VAR. This is so uniformly over the full forecast horizon and for all variables in the system. There is a tendency for the dominance of these methods to increase as the forecast horizon increases, again for all variables in the system.

(ii) For the cointegrated variables (Figs. 2(b) and (c)), the BVAR(lit) forecasts are poor relative to the other methods, especially for the first few periods ahead, where the forecast RMSE is 50–100% greater than that of the other methods. For these cointegrated variables, there is a clear advantage to using data-determined hyperparameters, as the BVAR(opt) model does, to allow for the effects of other variables in the system.

(iii) For the random-walk variable (Fig. 2(a)), the BVAR(opt) forecasts are marginally superior to the BVAR(lit) forecasts. This is explained by the fact that the data-determined hyperparameters allow for a choice that can shrink the coefficients closer to those of a random walk, and this tends to produce slightly better forecasts on average than those with the Litterman settings.

(iv) Overall, these figures show that there is a benefit to the use of data-determined procedures in forecasting. Not only is the use of a consistent cointegrating rank model selection method like PIC useful in forecasting from a reduced rank regression, but it is also clear that BVAR forecasts are improved by the use of data-determined selection of the hyperparameters.

4.4. *Policy analysis results*

The impulse response results are shown in Figs. 3(a)–(d). These figures graph the median (of the 1000 replications) impulse responses 1–30 periods out for the BVAR, RRR and ECM models against the true impulse responses (the solid line in each figure, as given in Eq. (7)). The median (rather than mean) responses are used so that the results are less affected by occasional very large responses that occur in the simulations. The impulse responses calculated from the unrestricted VAR regression had so many large responses that the graphs cannot be shown on the same figures without so distorting the scale that the graphs for the other models are indistinguishable. Instead, Figs. 5(a) and (b) show the full sampling distributions of the OLS responses 5 and 10 periods out. The dispersion of these distributions is enormous and the figures clearly show how unreliable the impulse

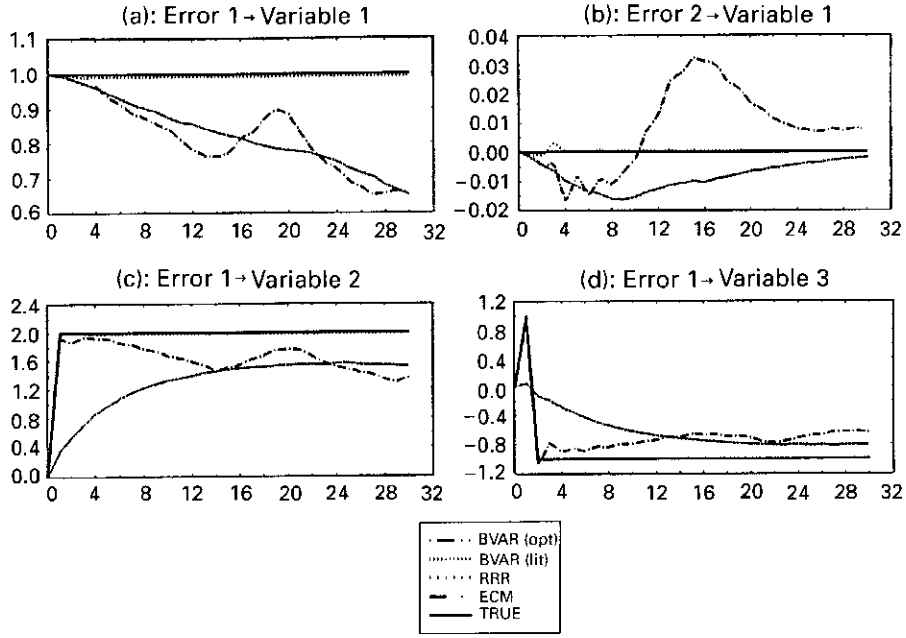

Fig. 3(a)–(d). Impulse responses.

responses calculated from an unrestricted VAR regression are, at least when there is a unit root in the system. We must conclude that in such cases the unrestricted OLS estimated impulse responses seem to be so unreliable that no meaningful inferences about policy effects can be drawn from them.

The results shown in Figs. 3(a)–(d) for the other methods of estimation seem much more reasonable. The main points to emerge are as follows:

(i) The median ECM responses are highly accurate. This is explained by the fact that in the ECM model, the form of the cointegrating links between the variables and the presence of a unit root in the first equation of the model is part of the prior specification. (Note that the cointegrating coefficients and the stationary dynamics, including lag order, are estimated in the ECM system). Thus, it is apparent that accurate structural knowledge pays off handsomely in delivering highly accurate impulse responses.

(ii) The median RRR impulse responses are also very accurate. Again this is explained by the fact that the correct cointegrating rank is selected in a large number of the simulations. In those cases the fitted model correctly incorporates a single unit root, and the estimated impulse responses then exhibit the persistence of shocks to the first equation on the first and second

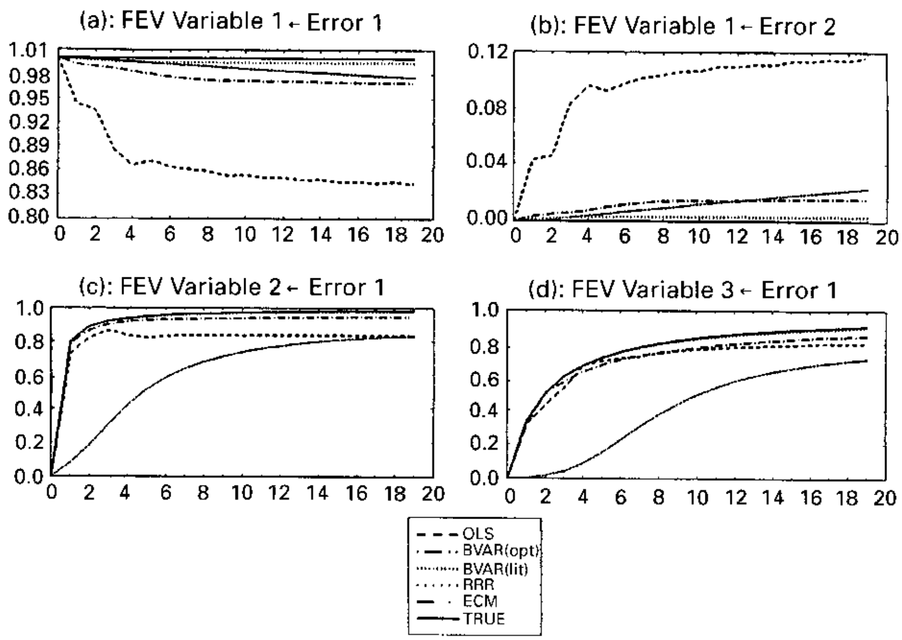

Fig. 4(a)–(d). FEV decompositions.

variables – see Figs. 3(a) and (c). Since the figures give median responses, they do not show that when the cointegrating rank is chosen incorrectly, the estimated impulse responses from the RRR tend to suffer from the same problem as those of an unrestricted VAR, viz. that some impulse response paths can be poorly estimated and even diverge if there is an explosive root. Nonetheless, the RRR responses are decidedly superior to the unrestricted VAR responses in general.

(iii) The median BVAR(opt) impulse responses appear to be more accurate for the first few periods than those of the BVAR(lit) – see Figs. 3(c) and (d) especially – but also seem to be more variable for the longer period responses. The most likely explanation of this phenomena is that the BVAR(opt) estimates are more influenced by cross equation effects because the data-determined tightness hyperparameter is generally larger than the Litterman setting (due to the presence of two cointegrating vectors in the true system).

Figs. 4(a)–(d) graph the mean simulated forecast error variance decompositions (FEVDs) against the true FEVDs as given in Eq. (8). As discussed earlier, for this system the FEVDs estimated by an unrestricted VAR are consistent because the random component in the forecast error variance is a scalar and scales

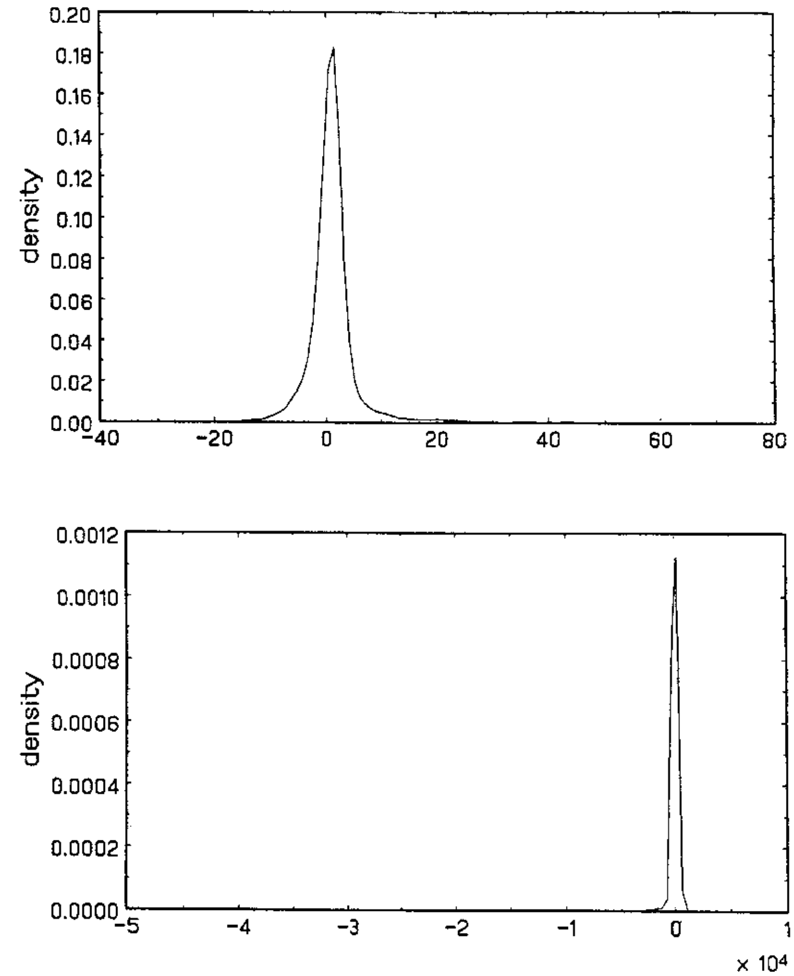

Fig. 5. Sampling distribution: (a) 5-period ahead OLS impulse responses (b) 10-period ahead OLS impulse responses.

out in the FEVD calculation – see (8) above. So this is a case where we may expect unrestricted VARs to perform satisfactorily. The main results can be summarized as follows.

(i) The unrestricted VAR estimated FEVDs display the most bias, especially with respect to the effect of shocks on the first variable.
(ii) The BVAR(lit) estimated FEVD's are also biased, especially with respect to the effect of the shocks in the first equation on the second and third variables.

(iii) The other methods generally seem to perform well in estimating the FEVDs, both in cases where there is shock persistence (Figs. 4(a), (c) and (d)) and where there is not (Fig. 4(b)).
(iv) In general, these simulations suggest that unrestricted VAR regression and BVAR(lit) regression are the least reliable methods for estimating FEVDs, at least when some unit roots and some cointegrated variables are present in the system.

Obviously, it is of interest to extend the simulations reported here to cases of different cointegrating rank, different cointegrating vector configurations, near unit roots and completely stationary systems. However, provided consistent model selection techniques that allow for the presence of some unit roots and cointegration are employed, it seems reasonable to expect that such automated methods of model-based policy analysis will generally perform better than unrestricted VAR regressions (see Fig. 5).

5. Conclusion

Unrestricted VARs have been extensively used in recent empirical research to assess the evidence in support of central propositions of macroeconomics, such as the role of money in the determination of aggregate output. Estimated impulse responses and forecast error decompositions have played a key role in these exercises. The calculation of long horizon impulse responses are now routine in this type of research and stem from the desire to learn about the long-run effects of shocks on the system. The approach has been vigorously pursued, for example, in studying the long-run effects of unanticipated monetary shocks on output, following the research of Sims (1980).

This paper raises some important issues about what we can expect to learn from this line of empirical research. Our asymptotic analysis shows that in nonstationary VAR models with some roots at or near unity the estimated impulse response matrices are inconsistent at long horizons and tend to random matrices rather than the true impulse responses. Thus, even in very large samples, we must inevitably expect uncertainty about policy analyses that are conducted using impulse responses that are estimated by unrestricted VARs. Our simulations indicate that there is also substantial sampling variation in these estimated responses in finite samples.

Some previous research (e.g. Spencer, 1989; Todd, 1990) has shown that estimated impulse responses and FEVDs can be very sensitive to changes in VAR model specification, such as the inclusion of trends and additional variables; and there has been debate about the robustness of the empirical findings in this line of research (see Todd, 1990, for an overview of the debate and some simulation analyses of sensitivity). Our results corroborate these earlier findings about unrestricted VAR impulse responses, give clear analytical reasons why impulse responses and FEVDs from unrestricted VARs are unreliable even in very large samples, and show that different models in the VAR class produce impulse responses and FEVDs with very different behaviour. Some models, like unrestricted VARs and Bayesian VARs produce inconsistent impulse responses and FEVDs. Others, like reduced rank regressions that employ consistent estimates of the cointegrating rank, and correctly specified error correction models produce consistent estimates of impulse responses and FEVDs. It is particularly important that the number of cointegrating relations in a system (and hence the number of unit roots) be estimated consistently. Model selection methods are important in achieving this. In particular, a reduced rank regression approach to impulse response analysis can be expected to improve upon unrestricted VARs only if the cointegrating rank selection methods work well in practice.

In general, our results echo the earlier findings of Christ (1975) for structural econometric models. While there certainly are differences in forecasting performance in linear time-series models, the most serious disagreements between time-series models arise in policy analyses. Our main conclusion is that differing treatments of nonstationarity in the models plays a big role in affecting the outcomes of policy analysis. Although this issue was not investigated by Christ, it seems likely (by analogy to our results for reduced rank regressions and error correction models) that similar effects to those we have discovered come into play in structural econometric models when unit roots or near unit roots are estimated.

Appendix A. Model formulations and proofs

A.1. The $I(1)/I(0)$ VAR representation

Construct the orthogonal matrix $H=[\beta_\perp,\beta]$, and define $z_t=H'y_t$. The system (2) transforms to

$$z_t = Bz_{t-1} + FW_t + \eta_t \quad \text{with } B = H'AH,$$
$$F = H'[\Psi_1,\ldots,\Psi_{p-1}](I_{p-1}\otimes H), \quad \eta_t = H'\varepsilon_t, \tag{2'}$$

and where $w_t'=(\Delta y_{t-1}',\ldots,\Delta y_{t-p+1}')(I\otimes H)$ is the vector of transformed difference regressors. Partitioning z_t, η_t and F conformably with the partition of H, and noting that B has the explicit partitioned form

$$B=\begin{bmatrix} I_s & \beta_\perp'\alpha \\ 0 & I_r+\beta'\alpha \end{bmatrix},$$

we can write $(2')$ as

$$z_{1t} = z_{1t-1} + \beta_\perp'\alpha z_{2t-1} + F_1 w_t + \eta_{1t} = z_{1t-1} + u_{1t} \quad \text{say}$$
$$z_{2t} = (I_r + \beta'\alpha)z_{2t-1} + F_2 w_t + \eta_{2t}.$$

In this representation of the VAR system, z_{1t} is I(1), z_{2t} is I(0), there are s unit roots in the first subsystem, and the second subsystem is stationary.

Define $x_t = (z'_{2t-1}, w'_t)'$, the transformed stationary components in the system, and then

$$x_t = \begin{bmatrix} I & 0 \\ 0 & I \otimes H' \end{bmatrix} \begin{bmatrix} z_{2t-1} \\ \Delta y_{t-1} \\ \cdot \\ \Delta y_{t-p+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \otimes H' \end{bmatrix} \begin{bmatrix} \beta' y_{t-1} \\ \Delta y_{t-1} \\ \cdot \\ \Delta y_{t-p+1} \end{bmatrix}$$

$$= \begin{bmatrix} I & 0 \\ 0 & I \otimes H' \end{bmatrix} \xi_t \quad \text{say.}$$

A.2. Alternate companion forms

It will be helpful in subsequent derivations to use alternate companion forms of the VAR model (1) that correspond directly to the model in levels and differences – see Eq. (2) – and the model in partitioned I(1)/I(0) format. We start by transforming (3) and (4) into the companion form for the model (2). This can be accomplished using the matrices

$$K = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & -I & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ I & -I & \cdots & -I \end{bmatrix} \quad \text{and} \quad K^{-1} = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & -I & \cdots & 0 \\ \cdot & \cdots & \cdots & \cdot \\ 0 & \cdots & I & -I \end{bmatrix},$$

giving the new companion form coefficient matrix

$$D = K^{-1}CK = \begin{bmatrix} I + \alpha\beta' & \Psi_1 & \cdots & \Psi_{p-1} \\ \alpha\beta' & \Psi_1 & \cdots & \Psi_{p-1} \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdots & I & 0 \end{bmatrix}. \tag{A.1}$$

Now define the orthogonal matrix

$$G = I_p \otimes H = \begin{bmatrix} \beta_\perp & \beta & 0 \\ 0 & 0 & I_{p-1} \otimes H \end{bmatrix} = [G_\perp, G_x], \quad G'_\perp = [\beta'_\perp, 0],$$

which we use to transform the companion matrix D again so that it corresponds with model (2′) where the variables are partitioned into I(1) and I(0) components.

Specifically, the matrix $E = G'K^{-1}CKG = G'DG$ is the coefficient matrix in the companion form of (2′). Note that it has the same eigenvalues as C and can be written in the partitioned form

$$E = G'DG = \begin{bmatrix} I_s & \beta_\perp'\alpha & \overline{\Psi}_1 & \cdots & \overline{\Psi}_{p-1} \\ 0 & I_r + \beta'\alpha & & & \\ 0 & \beta_\perp'\alpha & \overline{\Psi}_1 & \cdots & \overline{\Psi}_{p-1} \\ 0 & \beta'\alpha & & & \\ 0 & I_m & \cdot & & 0 \\ \cdot & \cdot & \cdot & & 0 \\ 0 & \cdots & I_m & & 0 \end{bmatrix} = \begin{bmatrix} I_s & E_{12} \\ 0 & E_{22} \end{bmatrix}, \tag{A.2}$$

where

$$E_{12} = [\beta_\perp'\alpha \ddot{\Psi}_1 \cdots \ddot{\Psi}_{p-1}] \quad \text{and} \quad E_{22} = \begin{bmatrix} I_r + \beta'\alpha & \ddot{\Psi}_1 & \cdots & & \ddot{\Psi}_{p-1} \\ \beta_\perp'\alpha & \overline{\Psi}_1 & \cdots & & \overline{\Psi}_{p-1} \\ \beta'\alpha & & & & \\ I_m & 0 & \cdots & & 0 \\ \cdot & & \cdots & & \cdot \\ 0 & 0 & \cdots & I_m & 0 \end{bmatrix},$$

and where $\ddot{\Psi}_k = \beta_\perp'\Psi_k H$, $\ddot{\Psi}_{\underline{k}} = \beta'\Psi_k H$, and $\overline{\Psi}_k = H'\Psi_k H$. When we take powers of E we get

$$E^i = G'D^iG = \begin{bmatrix} I_s & E_{12}(I + \cdots + E_{22}^{i-1}) \\ 0 & E_{22}^i \end{bmatrix} \to \begin{bmatrix} I_s & E_{12}(I - E_{22})^{-1} \\ 0 & 0 \end{bmatrix}, \tag{A.3}$$

since E_{22} has stable roots. The impulse response matrices can be rewritten in terms of the new companion form involving the matrix E as follows:

$$\begin{aligned} \Theta_i &= M'C^iM = M'KD^iK^{-1}M \\ &= M'KG \begin{bmatrix} I_s & E_{12}(I + \cdots + E_{22}^{i-1}) \\ 0 & E_{22}^i \end{bmatrix} G'K^{-1}M. \end{aligned} \tag{A.4}$$

In the near integrated case under (c′) we have

$$\begin{aligned} \Theta_i &= M'C^iM \\ &= M'KG \begin{bmatrix} \exp(in^{-1}\Gamma) & \exp(in^{-1}\Gamma)E_{12}(I + \cdots + E_{22}^{i-1}) \\ O(n^{-1}) & E_{22}^i \end{bmatrix} G'K^{-1}M. \end{aligned}$$

A.3. Proof of Lemma 2.2

From (A.4), we have $\Theta_i \to \bar{\Theta}$, where

$$\bar{\Theta} = M'KG \begin{bmatrix} I_s & E_{12}(I-E_{22})^{-1} \\ 0 & 0 \end{bmatrix} G'K^{-1}M = [\beta_\perp, \beta, 0, \ldots, 0]$$

$$\times \begin{bmatrix} I_s & E_{12}(I-E_{22})^{-1} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \beta'_\perp \\ \beta' \\ H' \\ 0 \\ \cdot \\ 0 \end{bmatrix}$$

$$= \beta_\perp \beta'_\perp + \beta_\perp E_{12}(I-E_{22})^{-1} Q = \beta_\perp \beta'_E,$$

as required. In case of (c$'$) and $i = fn$, we get $\Theta_i \to \bar{\Theta}_\Gamma = \beta_\perp \exp(f\Gamma)\beta'_E$.

A.4. Proof of Theorem 2.3

When (1) is estimated by unrestricted least squares we can write the estimated impulse response matrices in a form that is similar to the representation (A.4) above for the true impulse response matrices, *viz.*

$$\hat{\Theta}_i = M'\hat{C}^i M = M'K\hat{D}^i K^{-1}M = M'KG\hat{E}^i G'K^{-1}M.$$

In this expression $\hat{E}$ is formed from the unrestricted OLS estimate of the coefficients in the system (2$'$). Specifically,

$$\hat{E} = \begin{bmatrix} \hat{B} & \hat{F}_1 & \cdots & \hat{F}_{p-2} & \hat{F}_{p-1} \\ \hat{B}-I & \hat{F}_1 & \cdots & \hat{F}_{p-2} & \hat{F}_{p-1} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}.$$

Now $\hat{E} \underset{p}{\longrightarrow} E$, and, for fixed i, $\hat{E}^i \underset{p}{\longrightarrow} E^i$ as $n \to \infty$. In this case, the estimated impulse responses $\hat{\Theta}_i$ are consistent, proving the first part of part (i) of the theorem. The limit given in part (ii) follows using Lemma 2.2, as shown below.

The limit distribution of $\hat{\Theta}_i$ for fixed i is obtained as follows. We take differentials of these impulse response estimates with respect to the coefficient estimates giving

$$\mathrm{d}\hat{\Theta}_i = M\,\mathrm{d}\hat{C}^i M = M'KG\,\mathrm{d}\hat{E}^i G'K^{-1}M = M'KG \sum_{k=0}^{i-1} \hat{E}^{i-1-k}\,\mathrm{d}\hat{E}\hat{E}^k G'K^{-1}M.$$

Note that

$$\mathrm{d}\hat{E} = \begin{bmatrix} I \\ I \\ 0 \\ \cdot \\ 0 \end{bmatrix} [\mathrm{d}\hat{B}, \mathrm{d}\hat{F}_1, \ldots, \mathrm{d}\hat{F}_{p-1}] = K^{-1}M[\mathrm{d}\hat{B}_1 \vdots \mathrm{d}\hat{B}_2, \mathrm{d}\hat{F}_1, \ldots, \mathrm{d}\hat{F}_{p-1}]$$

$$= K^{-1}M[\mathrm{d}\hat{B}_1 \vdots \mathrm{d}\hat{B}_x] \quad \text{say},$$

$$= K^{-1}G'GM[\mathrm{d}\hat{B}_1 \vdots \mathrm{d}\hat{B}_x] = K^{-1}G'MH[\mathrm{d}\hat{B}_1 \vdots \mathrm{d}\hat{B}_x] = K^{-1}G'M[\mathrm{d}\hat{A}, \mathrm{d}\hat{\Psi}]G$$

$$= G'K^{-1}M[\mathrm{d}\hat{A}, \mathrm{d}\hat{\Psi}]G,$$

in the original coordinates of system (2). The partition $[\mathrm{d}\hat{B}_1 \vdots \mathrm{d}\hat{B}_x]$ above corresponds to the nonstationary and stationary coefficients in the transformed system (2′). From Theorem 5.7 and Remark 5.8 of Phillips (1995), we have

$$n^{1/2}[\hat{A} - A \vdots \hat{\Psi} - \Psi] \Rightarrow \mathrm{N}(0, \Sigma_\varepsilon \otimes G_x \Sigma_{xx}^{-1} G_x') = \mathrm{N}(0, \Sigma_\varepsilon \otimes G_\xi \Sigma_{\xi\xi}^{-1} G_\xi'),$$

where, $\Sigma_{xx} = E(x_t x_t')$, $\Sigma_{\xi\xi} = E(\xi_t \xi_t')$ and $\xi_t = [y_{t-1}'\beta, \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}']'$ and

$$G_\xi = \begin{bmatrix} \beta & 0 \\ 0 & I_{m(p-1)} \end{bmatrix}.$$

It follows that $\mathrm{d}\hat{\Theta}_i = \sum_{k=0}^{i-1} \Theta_{i-1-k}[\mathrm{d}\hat{A}, \mathrm{d}\hat{\Psi}]G\hat{E}^k G'K^{-1}M = \sum_{k=0}^{i-1} \hat{\Theta}_{i-1-k}[\mathrm{d}\hat{A}, \mathrm{d}\hat{\Psi}] K^{-1}\hat{C}^k M$ and $n^{1/2}(\hat{\Theta}_i - \Theta_i) \Rightarrow \mathrm{N}(0, V_i)$, where $V_i = N_i(\Sigma_\varepsilon \otimes G_x \Sigma_{xx}^{-1} G_x')N_i' = N_i(\Sigma_\varepsilon \otimes G_\xi \Sigma_{\xi\xi}^{-1} G_\xi')N_i'$ and $N_i = \sum_{k=0}^{i-1} \Theta_{i-1-k} \otimes M'C'^k K'^{-1}$. This completes the proof of part (ii) of the theorem.

When $i = fn$, where f is a constant, the consistent limit for $\hat{E}^i$ is no longer valid. Instead, the asymptotic distribution of the nonstationary components of $\hat{B}$ figures in the limit, as we now demonstrate.

Working in the transformed system (2′), we note that the limit distribution of those components of $\hat{B}$ which relate to the I(1) elements of the system (viz. the first s columns of B) is given by

$$n(\hat{B}_1 - B_1) \Rightarrow \left(\int_0^1 \mathrm{d}S_\eta S_1' \right) \left(\int_0^1 S_1 S_1' \right)^{-1}, \tag{A.5}$$

where S_1 is vector Brownian motion with covariance matrix = lrvar(u_{1t}), and S_η is vector Brownian motion with covariance matrix = lrvar(η_t) see Phillips (1995, Theorem 5.5) for the derivation of (5). Note that S_η and S_1 are correlated Brownian motions because η_{1t} is a component of u_{1t}. From (3) we see that $B_1' = [I_s, 0] = [B_{11}, B_{12}]$, say, so that when we partition $\hat{B}$ in the same way as B

we get

$$n(\hat{B}_{11} - I_s) \Rightarrow \left(\int_0^1 dS_{\eta_1} S_1'\right)\left(\int_0^1 S_1 S_1'\right)^{-1} = U \quad \text{say,}\ \hat{B}_{12} = O_p(n^{-1}). \tag{A.6}$$

Thus,

$$\hat{E} = \begin{bmatrix} \hat{B}_{11} & \hat{E}_{12} \\ O_p(n^{-1}) & \hat{E}_{22} \end{bmatrix} \quad \text{and}$$

$$\hat{E}^i = \begin{bmatrix} \hat{B}_{11}^i + O_p(n^{-1}) & \sum_{k=0}^i \hat{B}_{11}^{i-k}\hat{E}_{12}\hat{E}_{22}^k + O_p(n^{-1}) \\ O_p(n^{-1}) & \hat{E}_{22}^i + O_p(n^{-1}) \end{bmatrix}. \tag{A.7}$$

The eigenvalues of $\hat{E}_{22}$ converge in probability to the eigenvalues of E_{22}, which are the stable roots of the system. Therefore, $\hat{E}_{22}^i$ converges in probability to a zero matrix as $n \to \infty$. On the other hand,

$$\hat{B}_{11}^i = [I_s + (\hat{B}_{11} - I_s)]^i = [I_s + n(\hat{B}_{11} - I_s)/n]^{fn} \Rightarrow \exp(fU) \tag{A.8}$$

as $n \to \infty$. Let $i^* \to \infty$ be such that $i^* n^{-1} \to 0$. Then,

$$\sum_{k=0}^{i} \hat{B}_{11}^{i-k}\hat{E}_{12}\hat{E}_{22}^k = \hat{B}_{11}^i \sum_{k \leqslant i^*} \hat{B}_{11}^{-k}\hat{E}_{12}\hat{E}_{22}^k + \sum_{k > i^*} \hat{B}_{11}^{i-k}\hat{E}_{12}\hat{E}_{22}^k. \tag{A.9}$$

Now for $k \leqslant i^*$, we have $\hat{B}^{11-k} \xrightarrow[p]{} I$, and the first term on the right-hand side of (A.9) converges weakly to $\exp(fV)E_{12}(I - E_{12})^{-1}$. The second term converges in probability to zero since the roots of E_{22} are stable, and $\hat{B}_{11}^{i-k} = O_p(1)$. Thus

$$\hat{E}^i \Rightarrow \begin{bmatrix} \exp(fU) & \exp(fU)E_{12}(I - E_{22})^{-1} \\ 0 & 0 \end{bmatrix} = E_V \quad \text{say,}$$

and so $\hat{\Theta}_i \Rightarrow M'KGE_V G'K^{-1}M = \beta_\perp \exp(fU)\beta_\perp' + \beta_\perp \exp(fU)E_{12}(I - E_{22})^{-1}Q$, as required for part (iii) of the theorem.

Part (ii) follows by noting that, in place of (A.8), when $in^{-1} \to 0$ as $n \to \infty$ we get $\hat{B}_{11}^i \xrightarrow[p]{} I$, and hence $\hat{E}^i \xrightarrow[p]{} E_0$. Then, $\hat{\Theta}_i \xrightarrow[p]{} \bar{\Theta}$, as required.

To prove part (iv) we note that when the model (1) has near unit roots rather than unit roots, i.e. when assumption (c′) holds in place of (c), then we have the new coefficient matrix $A = \beta_\perp \exp(n^{-1}\Gamma)\beta_\perp' + \beta\beta' + \alpha\beta'$ in (2) rather than $A = I + \alpha\beta'$. Then, in place of (A.5) and (A.6) above, we have the alternative limit theory

$$n(\hat{B}_1 - B_1) \Rightarrow \left(\int_0^1 dS_\eta J_\Gamma'\right)\left(\int_0^1 J_\Gamma J_\Gamma'\right)^{-1}, \tag{A.5′}$$

and

$$n(\hat{B}_{11} - B_{11}) \Rightarrow \left(\int_0^1 \mathrm{d}S_{\eta_1} J_\Gamma' \right) \left(\int_0^1 J_\Gamma J_\Gamma' \right)^{-1} = U_\Gamma \quad \text{say,} \tag{A.6$'$}$$

where $J_\Gamma(r) = \int_0^r \exp\{(r-s)\Gamma\} \, \mathrm{d}S_1$ is a vector diffusion process – see Phillips (1988) for similar derivations. Part (iv) now follows upon appropriate redefinition of U.

A.5. Proof of Theorem 2.9

We proceed in the same general way as the proof of Theorem 2.3. As before, we can write $\hat{\Theta}_1 = M'K\hat{D}^i K^{-1} M$ and now

$$\hat{D} = \begin{bmatrix} I + \hat{\alpha}\hat{\beta}' & \hat{\Psi}_1 & \cdots & \hat{\Psi}_{p-1} \\ \hat{\alpha}\hat{\beta}' & \hat{\Psi}_1 & \cdots & \hat{\Psi}_{p-1} \\ 0 & I & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdots & I & 0 \end{bmatrix},$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the reduced rank regression estimates of α and β. We can take it here that β is normalized in such a way that its components are identified. Thus, we can set $\beta' = [I_r, -A_\beta]$, and require $\hat{\beta}$ to be normalized in the same way. These matrices can subsequently be orthonormalized by using the transformation $\beta \to \beta(\beta'\beta)^{-1/2}$. In the same way, we can define the coordinate system for the orthogonal complement space in such a way that $\beta_\perp' = [A_\beta', I_{m-r}]$ and orthonormalize the matrix with the transformation $\beta_\perp \to \beta_\perp(\beta_\perp'\beta_\perp)^{-1/2}$, doing the same thing for $\hat{\beta}_\perp$. Then, $\hat{A}_\beta \underset{p}{\longrightarrow} A_\beta$, and, in consequence, $\hat{\beta} \underset{p}{\longrightarrow} \beta$ and $\hat{\beta}_\perp \underset{p}{\longrightarrow} \beta_\perp$.

As in the proof of Theorem 2.3, we find that $\mathrm{d}\hat{\Theta}_i = \sum_{k=0}^{i-1} \hat{\Theta}_{i-1-k} [\mathrm{d}\hat{A}, \mathrm{d}\hat{\Psi}] K^{-1} \hat{C}^k M$, but now $\mathrm{d}\hat{A} = \mathrm{d}(\hat{\alpha}\hat{\beta}') = \mathrm{d}\hat{\alpha}\hat{\beta}' + \hat{\alpha}\,\mathrm{d}\hat{\beta}'$. Since $\hat{\beta}$ is $O(n^{-1})$ – consistent for β (recall that the components of β are identified), the dominant contribution to the limit distribution comes from $[\mathrm{d}\hat{\alpha}\hat{\beta}', \mathrm{d}\hat{\Psi}]$. If we define $\Phi = [\alpha, \Psi]$, then the asymptotic distribution of the reduced rank regression estimator of Φ is (see Ahn and Reinsel, 1990, Theorem 2)

$$n^{1/2}(\hat{\Phi} - \Phi) \Rightarrow \mathrm{N}(0, \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{\xi\xi}^{-1}),$$

where $\Sigma_{\xi\xi} = E(\xi_t \xi_t')$ and $\xi_t = [y_{t-1}'\beta, \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}']'$ as before. Next, write $\mathrm{d}\hat{\Phi} = [\mathrm{d}\hat{\alpha}, \mathrm{d}\hat{\Psi}]$ and then

$$n^{1/2}[(\hat{\alpha} - \alpha)\hat{\beta}', \hat{\Psi} - \Psi] \Rightarrow \mathrm{N}(0, \Sigma_{\varepsilon\varepsilon} \otimes G_\xi \Sigma_{\xi\xi}^{-1} G_\xi'). \tag{A.10}$$

It follows that

$$n^{1/2}(\hat{\Theta}_i - \Theta_i) \Rightarrow \mathrm{N}(0, V_i),$$

where $V_i = N_i(\Sigma_{\varepsilon\varepsilon} \otimes G_\xi \Sigma_{\xi\xi}^{-1} G_\xi') N_i'$ and N_i has the same form as before. This proves part (i) of the theorem.

Define $\hat{H} = [\hat{\beta}_\perp, \hat{\beta}]$ and $\hat{G} = I_p \otimes \hat{H}$. Let

$$\breve{E} = \hat{G}'\hat{D}\hat{G} = \begin{bmatrix} I_s & \hat{\beta}_\perp'\hat{\alpha} & \breve{\Psi}_1 & \cdots & \breve{\Psi}_{p-1} \\ 0 & I_r + \hat{\beta}'\hat{\alpha} & & & \\ 0 & \hat{\beta}_\perp'\hat{\alpha} & \breve{\Psi}_1 & \cdots & \breve{\Psi}_{p-1} \\ 0 & I_r + \hat{\beta}'\hat{\alpha} & & & \\ 0 & I_m & . & . & 0 \\ . & . & . & & 0 \\ 0 & \cdots & I_m & & 0 \end{bmatrix} = \begin{bmatrix} I_s & \breve{E}_{12} \\ 0 & \breve{E}_{22} \end{bmatrix} \quad \text{say,}$$

where $\breve{\Psi}_k = \hat{H}'\hat{\Psi}_k\hat{H}$, $k = 1, \ldots, p-1$. Then

$$\breve{E}^i = \hat{G}^i\hat{D}^i\hat{G} = \begin{bmatrix} I_s & \breve{E}_{12}(I + \breve{E}_{22} + \ldots \breve{E}_{22}^{i-1}) \\ 0 & \breve{E}_{22}^i \end{bmatrix} \underset{p}{\longrightarrow} \begin{bmatrix} I_s & E_{12}(I - E_{22})^{-1} \\ 0 & 0 \end{bmatrix}.$$

It follows that

$$\begin{aligned} \hat{\Theta}_i &= M'K\hat{D}^iK^{-1}M \\ &= M'K\hat{G}\breve{E}^i\hat{G}'K^{-1}M \underset{p}{\longrightarrow} M'KG \begin{bmatrix} I_s & E_{12}(I - E_{22})^{-1} \\ 0 & 0 \end{bmatrix} G'K^{-1}M \\ &= \beta_\perp\beta_\perp' + \beta_\perp E_{12}(I - E_{22})^{-1}Q = \bar{\Theta}, \end{aligned}$$

giving the required result (ii).

A.6. Proof of Theorem 3.1.

Taking differentials of $\hat{F}$ we have $\mathrm{d}\hat{F} = \sum_{i=0}^{h-1}\{\mathrm{d}\hat{\Theta}_i\hat{\Sigma}\hat{\Theta}_i' + \hat{\Theta}_i\mathrm{d}\hat{\Sigma}\hat{\Theta}_i' + \hat{\Theta}_i\hat{\Sigma}\mathrm{d}\hat{\Theta}_i'\}$, and vectorizing yields

$$\begin{aligned} \mathrm{vec}(\mathrm{d}\hat{F}) &= \sum_{i=0}^{h-1} [\{(I \otimes \hat{\Theta}_i\hat{\Sigma}) + (\hat{\Theta}_i\hat{\Sigma} \otimes I)K_{mm}\}\mathrm{vec}(\mathrm{d}\hat{\Theta}_i) \\ &\quad + (\hat{\Theta}_i \otimes \hat{\Theta}_i)\mathrm{vec}(\mathrm{d}\hat{\Sigma})] \\ &= \sum_{i=0}^{h-1} [\{(I \otimes \hat{\Theta}_i\hat{\Sigma}) + (\hat{\Theta}_i\hat{\Sigma} \otimes I)K_{mm}\}\mathrm{vec}(\mathrm{d}\hat{\Theta}_i) + (\hat{\Theta}_i \otimes \hat{\Theta}_i)D\mathrm{d}\hat{\sigma}] \\ &= \sum_{i=0}^{h-1} [\{(I \otimes \hat{\Theta}_i\hat{\Sigma}) + (\hat{\Theta}_i\hat{\Sigma} \otimes I)K_{mm}\}\hat{N}_i\mathrm{vec}(\mathrm{d}[\hat{A}, \hat{\Psi}]) \\ &\quad + (\hat{\Theta}_i \otimes \hat{\Theta}_i)D\mathrm{d}\hat{\sigma}]. \end{aligned}$$

Here $\hat{A}$ and $\hat{\Psi}$ are the OLS estimates of the coefficient matrices in (2), $\hat{N}_i = \sum_{k=0}^{i-1} \hat{\Theta}_{i-1-k} \otimes M'\hat{C}'^k K'^{-1}$ and $\text{vec}(\Sigma) = D\sigma$, σ is the vector of nonredundant elements of Σ and D is the duplication matrix. The limit distribution of the coefficient matrices and the covariance matrix estimates are independent and are given by

$$n^{1/2}[\hat{A} - A, \hat{\Psi} - \Psi] = n^{1/2}[(\hat{\alpha} - \alpha)\hat{\beta}', \hat{\Psi} - \Psi] \Rightarrow \text{N}(0, \Sigma_{\varepsilon\varepsilon} \otimes G_\xi \Sigma_{\xi\xi}^{-1} G_\xi'),$$

$$n^{1/2}(\hat{\sigma} - \sigma) \Rightarrow \text{N}(0, D^+ \text{var}(\varepsilon_t \otimes \varepsilon_t) D^{+\prime}),$$

where $D^+ = (D'D)^{-1}D'$ is a generalized inverse of D. If the errors in (2) are normally distributed, then $\text{var}(\varepsilon_t \otimes \varepsilon_t) = 2P_D(\Sigma_\varepsilon \otimes \Sigma_\varepsilon)$, where $P_D = D(D'D)^{-1}D'$, and the covariance matrix in the second limit distribution is simply $2D^+(\Sigma_\varepsilon \otimes \Sigma_\varepsilon) D^{+\prime}$. Part (i) of the theorem follows directly.

To prove part (ii), we first write $h^{-1}\hat{F}(h) = h^{-1}\sum_{i=0}^{h-1} \hat{\Theta}_i \hat{\Sigma}_\varepsilon \hat{\Theta}_i' = f^{-1}\sum_{i=0}^{h-1} \int_{(i-1)/n}^{i/n} \hat{\Theta}_i \hat{\Sigma}_\varepsilon \hat{\Theta}_i' \, \text{d}s$. From the proof of Theorem 2.3, $\hat{\Theta}_i = M'KG\hat{E}^i G'K^{-1}M$, and for $i = ns$, we have as in (A.7)–(A.9)

$$\hat{E}^i = \begin{bmatrix} \hat{B}_{11}^i + \text{O}_\text{p}(n^{-1}) & \hat{B}_{11}^i \hat{E}_{12}(I - \hat{E}_{22})^{-1} + \text{o}_\text{p}(1) \\ \text{O}_\text{p}(n^{-1}) & \hat{E}_{22}^i + \text{O}_\text{p}(n^{-1}) \end{bmatrix}, \qquad \hat{B}_{11} \Rightarrow \exp(sU)$$

and $\hat{\Theta}_i \Rightarrow \beta_\perp \exp(sU)\beta_E'$. It follows that

$$h^{-1}\hat{F}(h) = f^{-1}\sum_{i=0}^{h-1} \int_{(i-1)/n}^{i/n} \hat{\Theta}_i \hat{\Sigma}_\varepsilon \hat{\Theta}_i' \text{d}s \Rightarrow f^{-1}\int_0^f \beta_\perp \exp(sU)\beta_E' \Sigma_\varepsilon \beta_E \times \exp(sU')\beta_\perp' \, \text{d}s,$$

giving the stated result (a). Result (b) follows in a similar way. We have

$$\hat{\tau}_{jk,h} = h^{-1}\sum_{i=0}^{h-1} \hat{\varphi}_{i,jk}^2 = h^{-1}\sum_{i=0}^{h-1} (\hat{\Theta}_i \hat{P})_{jk}^2$$

$$= f^{-1}\sum_{i=0}^{h-1} \int_{(i-1)/n}^{i/n} (\hat{\Theta}_i \hat{P})_{jk}^2 \, \text{d}s \Rightarrow f^{-1}\int_0^f (\beta_\perp \text{e}^{sU} \beta_E' P)_{jk}^2 \, \text{d}s,$$

giving the stated result. Part (c) is a direct consequence of part (b). Part (iii) follows in a straightforward way using the near integrated asymptotics, and then the random matrix U is replaced with U_Γ in the above formulae.

A.7. Proof of Theorem 3.4

From (5) the optimal predictor is $y_{n,h} = \sum_{i=h}^{n+h-1} \Theta_i \varepsilon_{n+h-i}$. When $h = fn$ and $n \to \infty$, we deduce from Lemma 2.2 that

$$n^{-1/2} y_{n,h} \sim \Theta_h n^{-1/2} \sum_{i=1}^n \varepsilon_i \Rightarrow \bar{\Theta} S(1) = \beta_\perp \beta_E' S(1),$$

as required for part (i). In a similar way, since $\hat{\Theta}_h \Rightarrow \beta_\perp \exp(fU)\beta_E'$, we obtain part (ii). Parts (iii) and (iv) are immediate.

A.8. Proof of Theorem 3.7

The proof of part (i) is the same as that of Theorem 3.2 (i) in all key respects. However, RRR rather than OLS estimates of the coefficients in (2) are used. We then have $[\hat{A}, \hat{\Psi}] = [\hat{\alpha}\hat{\beta}', \hat{\Psi}]$, and we use the limit theory (A.10) for the RRR coefficient estimates. The stated result follows in the same way as Theorem 3.2(i). Part (ii) of the theorem is a consequence of the consistency of the RRR impulse responses for long horizons that was shown in Theorem 2.9(ii).

A.9. Proof of Theorem 3.9

The stated results follow in the same way as Theorem 3.4, but rely on the consistency of the RRR impulse responses.

References

Ahn, S.K., Reinsel, G.C., 1990. Estimation for partially nonstationary multivariate autoregressive models. Journal of the American Statistical Association 85, 813–823.

Blanchard, O.J., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. American Economic Review 79, 655–673.

Chao, J., Phillips, P.C.B., 1994. Bayesian model selection in partially non-stationary vector autoregressive processes with reduced rank structure. Yale University, Mimeographed, .

Christ, C., 1975. Judging the performance of econometric models of the US Economy. International Economic Review 16, 54–74.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

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

Lutkepohl, H., 1993. Introduction to Multiple Time Series Analysis, 2nd ed. Springer, New York.

Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: Part 1. Econometric Theory 4, 468–497.

Park, J.Y., Phillips, P.C.B., 1989. Statistical inference in regressions with integrated processes: Part 2. Econometric Theory 5, 95–131.

Phillips, P.C.B., 1988a. Multiple regression with integrated processes. In: Prabhu, N.U. (Ed.), Statistical Inference from Stochastic Processes, Contemporary Mathematics, vol. 80, pp. 79–106.

Phillips, P.C.B., 1988b. Regression theory for near-integrated time series. Econometrica 56, 1021–1044.

Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. Econometrica 59, 283–306.

Phillips, P.C.B., 1992. Bayes methods for trending multiple time series with an empirical application to the U.S. economy. Cowles foundation discussion paper, No. 1025.

Phillips, P.C.B., 1994a. Model determination and macroeconomic activity. Cowles foundation discussion paper, No. 1083.

Phillips, P.C.B., 1994b. Bayes models and forecasts of Australian macroeconomic time series. In: Hargreaves, C.P. (Ed.), Nonstationary Time Series and Cointegration. Oxford University Press, Oxford, pp. 53–86.

Phillips, P.C.B., 1995a. Bayesian model selection and prediction with empirical applications. Journal of Econometrics, in press.

Phillips, P.C.B., 1995. Fully modified least squares and vector autoregression. Econometrica 63, 1023–1078.

Phillips, P.C.B., Durlauf, S.N., 1986. Multiple time series with integrated variables. Review of Economic Studies 53, 473–496.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. Review of Economic Studies 57, 99–125.

Sims, C.A., 1980. Comparisons of inter-war and post-war business cycles: monetarism reconsidered. American Economic Review 70, 250–257.

Sims, C.A., Stock, J.H., Watson, M.W., 1990. Inference in linear time series models with some unit roots. Econometrica 58, 113–144.

Sims, C.A., Zha, T., 1994. Error bands for impulse responses. Mimeographed, Yale University.

Spencer, D.E., 1989. Does money matter? The robustness of evidence from vector autoregressions. Journal of Money Credit and Banking 21, 442–454.

Todd, R.M., 1990. Vector autoregression evidence on monetarism: another look at the robustness debate. Federal Reserve Bank of Minneapolis Quarterly Review, 19–37.