ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

VAR Modeling and Business Cycle Analysis: A Taxonomy of Errors

D.S. Poskitt and Wenying Yao

April 2012

Working Paper 11/12

VAR Modeling and Business Cycle Analysis: A Taxonomy of Errors

D. S. Poskitt*and Wenying Yao†

April 19, 2012

Department of Econometrics and Business Statistics
Monash University
VIC 3800, Australia

Abstract

In this article we investigate the theoretical behaviour of finite lag *VAR(n)* models fitted to time series that in truth come from an infinite order $VAR(\infty)$ data generating mechanism. We show that overall error can be broken down into two basic components, an estimation error that stems from the difference between the parameter estimates and their population ensemble *VAR(n)* counterparts, and an approximation error that stems from the difference between the *VAR(n)* and the true $VAR(\infty)$. The two sources of error are shown to be present in other performance indicators previously employed in the literature to characterize, so called, truncation effects. Our theoretical analysis indicates that the magnitude of the estimation error exceeds that of the approximation error, but experimental results based upon a prototypical real business cycle model indicate that in practice the approximation error approaches its asymptotic position far more slowly than does the estimation error, their relative orders of magnitude notwithstanding. The experimental results suggest that with sample sizes and lag lengths like those commonly employed in practice *VAR(n)* models are likely to exhibit serious errors of both types when attempting to replicate the dynamics of the true underlying process and that inferences based on *VAR(n)* models can be very untrustworthy.

Keywords: VAR, estimation error, approximation error, RBC model.

JEL classification: C18, C32, C52, C54, E37

*Corresponding author: D.S. Poskitt, Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia, Donald.Poskitt@monash.edu.au, Tel: +61 3 99059378, Fax: +61 3 99055474.

†Department of Econometrics and Business Statistics, Monash University.

1 Introduction

Following on from the pioneering work of Sims (1980) on the relationship between abstract macroeconomic variables and stylized facts as represented by statistical time series, vector autoregressive (VAR) models have become the workhorse of much macroeconomic modeling. In structural VAR ($SVAR$) models, for example, VARs coupled with restrictions derived from economic theory are used to examine the effects of structural shocks on key macroeconomic variables, see Christiano, Eichenbaum, and Vigfusson (2006) and Kascha and Mertens (2009) for recent contributions. In dynamic stochastic general equilibrium (DSGE) models, VARs are used as auxiliary models for indirect estimation of the DSGE parameters (Smith, 1993), and to provide approximations to the solutions of DSGE models that have been expanded around their steady state (DelNegro and Schorfheide, 2004).

Although VARs are widely used, recent research has given rise to considerable debate about their usefulness. Chari, Kehoe, and McGrattan (2007) examine a stylized business cycle model and find that the impulse response function (IRF) computed from a finite order VAR yields a poor characterization of the true responses. Similarly, in a study of a real business cycle (RBC) model Erceg, Guerrieri, and Gust (2005) conclude that the error associated with using a finite order VAR model can be large and attribute this to small–sample error. In contrast, Ravenna (2007) shows that a finite order $SVAR$ model can indeed lead to inaccurate estimates of the true IRFs but indicates that this may not be a small–sample problem. Ravenna (2007) suggests that the error derives from two separate sources: a "truncation bias" and an "identification bias" .

In the above papers Erceg, Guerrieri, and Gust (2005), Chari, Kehoe, and McGrattan (2007) and Ravenna (2007) recognize, of course, that DSGE models with linear transition laws and quadratic preferences, and linear or log linear approximations to the equilibrium decision rules of nonlinear RBC models, lead to state space structures that are observationally equivalent to vector autoregressive moving average ($VARMA$) processes and therefore that the root of the problem lies in the fact that the true underlying data generating mechanism is a $VAR(\infty)$ process. Consider a k-dimensional, zero mean, stationary $VARMA$ process

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + u_t + \Theta_1 u_{t-1} + \cdots + \Theta_q u_{t-q}, \tag{1.1}$$

where u_t is a k-dimensional zero mean reduced form white noise stochastic disturbance with variance-covariance matrix Σ. If we denote the autoregressive and moving average polynomials by $\Phi(z) = I - \Phi_1 z - \cdots - \Phi_p z^p$ and $\Theta(z) = I + \Theta_1 z + \cdots + \Theta_q z^q$ respectively, then the VAR representation of the process is the $VAR(\infty)$

$$Y_t = \Psi_1 Y_{t-1} + \Psi_2 Y_{t-2} \cdots + u_t = \sum_{i=1}^{\infty} \Psi_i Y_{t-i} + u_t \tag{1.2}$$

where $\Psi(z) = I - \sum_{i=1}^{\infty} \Psi_i z^i = \Theta(z)^{-1}\Phi(z)$. For ease of exposition we will suppose that

$\det \Phi(z) \neq 0$ and $\det \Theta(z) \neq 0$, $|z| \leq 1$, namely, that Y_t is stationary and invertible and that u_t corresponds to the fundamental innovations process.[1]

In situations where the theoretical background gives rise to the model in (1.1) it might be conjectured – on the basis of Weirstrass' approximation theorem – that a *VAR* of high order can be used to approximate the true *VARMA* structure reasonably well. Results in the recent literature suggest, however, that such a result may have little practical relevance. Chari, Kehoe, and McGrattan (2007), for example, conclude from their analysis that the currently available data is prohibitive, leading to *VAR*s that have too short a lag length and that provide poor approximations and unreliable inferences. For a simulated model that has both DSGE elements and data dynamics Kapetanios, Pagan, and Scott (2007) suggest that a sample of 30,000 observations with a *VAR* of order 50 is required to adequately capture the effect of some of the structural shocks. Ravenna (2007) also points out that using a *VAR* to characterize the dynamics of a model that in truth leads to a *VARMA* structure can be misleading, and warns researchers to be cautious when relying on evidence from *VAR*s to build such models.

The purpose of this paper is to provide a detailed theoretical examination of the loss incurred when approximating a $VAR(\infty)$ process by a finite lag *VAR(n)* model. We present the limiting properties of the estimated coefficients of a fitted *VAR(n)* model and show how these translate into other statistics of interest such as impulse responses and forecast error variances. This process leads to a classification of error into two distinct parts. An estimation error that stems from the difference between the estimated *VAR(n)* and its theoretical counterpart, and an approximation error that comes from the difference between the theoretical minimum mean squared error (MMSE) *VAR(n)* approximation and the true $VAR(\infty)$ process. Our results show that the estimation error, a stochastic quantity that varies along different sample paths, and the approximation error, a deterministic quantity that is fixed by the process parameters, are fundamental sources of error that are present in other performance measures employed in the literature.

Given knowledge of the true representation of the theoretical data generating mechanism, the magnitude of the estimation error and the approximation error can be computed from the population ensemble parameters. Application of such an analysis to the classical RBC model of Hansen (1985) indicates that in practice the approximation error can be very large and can dominate the estimation error, even when the sample size is extremely large. Indeed, our results suggest that the common rationale for using a *VAR(n)* model – that a *VAR(n)* will approximate a $VAR(\infty)$ process well in the limit, that is, as $n \to \infty$ as $T \to \infty$ – may not be justified unless T is enormous. Faced with sample sizes and lag lengths like those commonly used in practice, our results also suggest that differences in the estimation error and approximation error from one process to the next may go some way in explaining and reconciling the different results and conclusions seen hitherto in the literature.

The remainder of the paper is organized as follows. The estimation error and approximation

[1] See Fernndez-Villaverde, Rubio-Ramandiacute;rez, Sargent, and Watson (2007) and Lippi and Reichlin (1994) for a discussion of issues associated with non-invertibility and non-fundamental innovations.

error of a *VAR(n)* model are defined and analyzed in Sections 2 and 3, and the role they play in determining the "truncation bias" and "identification bias" in *SVAR*s is discussed in Section 4. Section 5 demonstrates the operation of our results in the context of a classical RBC model. Section 6 presents a brief conclusion. An appendix presents the algorithms employed to calculate the population ensemble parameters needed in Section 5.

Throughout the paper C, $0 < C < \infty$, will stand for a positive constant, not necessarily the same one, and $\lfloor x \rfloor$ will denote the integer part of x. For any $r \times c$ matrix M we will employ the standard notation $\|M\| = \sqrt{\sum_{i=1}^{r} \sum_{j=1}^{c} m_{ij}^2}$ for the Euclidean norm where $m_{ij} = [M]_{ij}$ is the ijth element of M. Let $\mathcal{L}^2$ be the complete Hilbert space of square integrable matrix functions P with support $(-\pi, \pi]$ and norm $\|P\| = \langle P, P \rangle^{1/2}$ where the inner product $\langle P_1, P_2 \rangle = \operatorname{tr} \int_{-\pi}^{\pi} P_1 P_2^* d\omega$. We will use $\|P(z)\| = \|P(e^{\imath\omega})\|$ to denote the norm of any matrix polynomial or power series $P(z) = \sum_{i \geq 0} P_i z^i$ such that $P(e^{\imath\omega}) \in \mathcal{L}^2$.

2 VAR Modeling and Estimation Error

Let

$$Y_t = \hat{\Psi}_{n1} Y_{t-1} + \hat{\Psi}_{n2} Y_{t-2} \cdots + \hat{\Psi}_{nn} Y_{t-n} + \hat{u}_{nt}, \tag{2.1}$$

denote a *VAR(n)* model fitted to a realization $Y_1, \ldots, Y_T$ on a wide sense stationary process where the estimated autoregressive coefficient matrices $\hat{\Psi}_{ni}$, $i = 1, \ldots, n$, are calculated using the empirical Yule–Walker equations, least squares, or the multivariate Burg algorithm. The population ensemble counterpart to (2.1) is

$$Y_t = \Psi_{n1} Y_{t-1} + \Psi_{n2} Y_{t-2} \cdots + \Psi_{nn} Y_{t-n} + u_{nt}, \tag{2.2}$$

wherein the coefficients Ψ_{ni}, $i = 1, 2, \cdots, n$, are obtained by solving the n-th order block Toeplitz (Yule–Walker) equations

$$\Gamma(h) - \Psi_{n1}\Gamma(h-1) - \cdots - \Psi_{nn}\Gamma(h-n) = \delta_{0h}\Sigma_n, \quad h = 0, 1, \ldots, n. \tag{2.3}$$

where δ_{jk} denotes the so–called Kronecker delta, taking the value one when $j = k$ and zero otherwise. In equation (2.3) the autocovariance sequence $\Gamma(h) = \mathbb{E}(Y_t Y_{t-h}')$, $h = 0, \pm 1, \pm 2, \cdots$, and $\Sigma_n = \mathbb{E}(u_{nt} u_{nt}')$, the MMSE from the prediction of Y_t from n previous values. The link between (2.1) and (2.2) comes from observing that under appropriate regularity the three estimators are asymptotically equivalent (Poskitt, 1994) and $\lim_{T\to\infty} \hat{\Psi}_{ni} = \Psi_{ni}$, $i = 1, 2, \cdots, n$, almost surely. We will state this result as a formal theorem (See Hannan and Deistler, 1988, Theorem 7.4.5)

Theorem 1 *Suppose that Y_t admits the autoregressive representation $Y_t = \sum_{i=1}^{\infty} \Psi_i Y_{t-i} + u_t$, as in (1.2), where u_t is a stationary martingale difference process with $\mathbb{E}(u_t|\mathfrak{F}_{t-1}) = 0$ and $\mathbb{E}(u_t u_t'|\mathfrak{F}_{t-1}) = \Sigma$, with $\mathfrak{F}_t$ denoting the σ–algebra generated by u_s, $s \leq t$. Set $\Psi(z) = I -$*

$\sum_{i=1}^{\infty}\Psi_i z^i$ *and require that* $det\,\Psi(z)\neq 0$, $|z|\leq 1$, $\sum_{i=1}^{\infty} i^{1/2}\|\Psi_i\|<\infty$ *and that* u_t *has finite fourth moments. Then for* $n\leq N_T=o((T/\log T)^{1/2})$ *there exists a* C, $0<C<\infty$, *such that*

$$\max_{1\leq i\leq n}\|\hat{\Psi}_{ni}-\Psi_{ni}\|\leq(\log T/T)^{1/2}C$$

uniformly in n.

A significant feature of Theorem 1 is that as the sample size T increases the *VAR(n)* coefficient estimates do not converge to the first n autoregressive coefficients of the $VAR(\infty)$ representation of the true process, but to the coefficient matrices of the optimal linear predictor of Y_t based on $Y_{t-1},\ldots,Y_{t-n}$. In the DSGE and RBC literature it is commonly asserted that $\hat{\Psi}_{n1},\cdots,\hat{\Psi}_{nn}$ provide consistent estimates of $\Psi_1,\cdots,\Psi_n$ (see Ravenna, 2007, p.2057 for example), and consequences arising from the use of $\hat{\Psi}_n(z)=I-\sum_{i=1}^{n}\hat{\Psi}_{ni}z^i$ to estimate $\Psi(z)$ are often referred to as truncation effects. However, at the risk of getting ahead of ourselves, we note that although the statement that $\hat{\Psi}_n(z)$ yields a consistent estimate of $\Psi(z)$ is formally correct, in a sense made clear immediately below, the assertion is *per se* an oversimplification that fails to recognize that detailed particulars and aspects of the technical complexity in the proof can be seen in observed behaviour.

The trivial decomposition $\hat{\Psi}_n(z)-\Psi(z)=\{\hat{\Psi}_n(z)-\Psi_n(z)\}+\{\Psi_n(z)-\Psi(z)\}$ where $\Psi_n(z)=I-\sum_{i=1}^{n}\Psi_{ni}z^i$ indicates that truncation error is made up of two components, an estimation error $\hat{\Psi}_n(z)-\Psi_n(z)$ and an approximation error $\Psi_n(z)-\Psi(z)$. Applying the triangular inequality gives us

$$\|\hat{\Psi}_n(z)-\Psi(z)\|\leq\|\hat{\Psi}_n(z)-\Psi_n(z)\|+\|\Psi_n(z)-\Psi(z)\|\,.$$

An immediate consequence of Theorem 1 is that the estimation error satisfies

$$\|\hat{\Psi}_n(z)-\Psi_n(z)\|=\|\sum_{i=1}^{n}(\Psi_{ni}-\hat{\Psi}_{ni})z^i\|=O\{n(\log T/T)^{1/2}\}\,. \tag{2.4}$$

For the approximation error we have

$$\begin{aligned}\|\Psi_n(z)-\Psi(z)\| &\leq \|\sum_{i=1}^{n}(\Psi_i-\Psi_{ni})z^i\|+\|\sum_{i=n+1}^{\infty}\Psi_i z^i\|\,,\\ &\leq C\sum_{i=n+1}^{\infty}\|\Psi_i\|\,,\quad 0<C<\infty,\end{aligned} \tag{2.5}$$

where the last line in (2.5) follows from Baxter's inequality Baxter (1962). Since by assumption $n\leq N_T=o((T/\log T)^{1/2})$ and $\sum_{i=1}^{\infty}i^{1/2}\|\Psi_i\|<\infty$ it follows from (2.4) and (2.5) that the overall truncation error $\|\hat{\Psi}_n(z)-\Psi(z)\|$ will converge to zero asymptotically. But in passage to the limit the estimation error and the approximation error converge at different rates and, as will be illustrated below, these differences are reflected in observed behaviour.

3 VAR(n) Approximation Error

From (2.5) it is apparent that the magnitude of the approximation error $\|\Psi_n(z)-\Psi(z)\|$ depends on the rate of decay of the $VAR(\infty)$ coefficients and the order n of the *VAR(n)* approximation. Under the conditions of Theorem 1 we have $n^{1/2}\|\Psi_n\| \to 0$ as $n \to \infty$ and $\sum_{i=n+1}^{\infty}\|\Psi_i\| = O(n^{-1/2})$, and we can therefore conclude that $\|\Psi_n(z) - \Psi(z)\| \leq Cn^{-1/2}$. In DSGE and RBC applications tighter bounds obtain. In such cases the true data generating mechanism is a stationary and invertible *VARMA* process, so there exists a complex constant ζ, $|\zeta| < 1$, such that $\|\Psi_n\| \leq C|\zeta|^n$, and hence $\sum_{i=n+1}^{\infty}\|\Psi_i\| \leq C|\zeta|^{n+1}/(1-|\zeta|)$ and the approximation error declines geometrically.

Now let us suppose that the criterion function

$$\rho_T(n) = \log\det\hat{\Sigma}_n + nk^2P_T/T\,, \quad P_T > 1\,, \quad n \leq N_T\,,$$

is used to determine the order of the fitted autoregression where, from the sample Yule–Walker equations, $\hat{\Sigma}_n = \hat{\Gamma}_T(0) - \sum_{i=1}^{n}\hat{\Psi}_{ni}\hat{\Gamma}_T(-i)$ and $\hat{\Gamma}_T(h) = \hat{\Gamma}_T(-h)' = T^{-1}\sum_{t=h+1}^{T}Y_tY_{t-h}'$. A second order Taylor series expansion of the logarithm of a determinant gives us

$$\log\det\hat{\Sigma}_n = \log\det\Sigma + \mathrm{tr}\{\Sigma^{-1}(\hat{\Sigma}_n - \Sigma)\} - \frac{1}{2}\|\Sigma^{-1}(\hat{\Sigma}_n - \Sigma)\|^2 + o(\|\hat{\Sigma}_n - \Sigma\|^2) \tag{3.1}$$

and substituting $\mathrm{tr}\{\Sigma^{-1}(\hat{\Sigma}_n - \Sigma)\} = \mathrm{tr}\{\Sigma^{-1}(\Sigma_n - \Sigma)\} + \mathrm{tr}\{\Sigma^{-1}(\hat{\Sigma}_n - \Sigma_n)\}$ into (3.1) yields

$$\begin{aligned}\log\det\hat{\Sigma}_n = \log\det\Sigma + \mathrm{tr}\{\Sigma^{-1}(\Sigma_n - \Sigma)\} + \\ \mathrm{tr}\{\Sigma^{-1}(\hat{\Sigma}_n - \Sigma_n)\} - \frac{1}{2}\|\Sigma^{-1}(\hat{\Sigma}_n - \Sigma)\|^2 + o(\|\hat{\Sigma}_n - \Sigma\|^2)\,.\end{aligned} \tag{3.2}$$

Employing Theorem 1 in conjunction with the bound $\|\hat{\Gamma}_T(h) - \Gamma(h)\| = O((\log T/T)^{1/2})$ (Hannan and Deistler, 1988, Theorem 7.4.3) it is straightforward to deduce from the expansion

$$\hat{\Sigma}_n - \Sigma_n = \{\hat{\Gamma}_T(0) - \Gamma(0)\} - \sum_{i=1}^{n}\hat{\Psi}_{ni}\{\hat{\Gamma}_T(-i) - \Gamma(-i)\} + \{\hat{\Psi}_{ni} - \Psi_{ni}\}\Gamma(-i)$$

that $\|\hat{\Sigma}_n - \Sigma_n\| = O(n(\log T/T)^{1/2})$. Now,

$$\Sigma_n - \Sigma = \int_{-\pi}^{\pi}\Psi_n(e^{\imath\omega})S_y(\omega)\Psi_n(e^{\imath\omega})^* - \Psi(e^{\imath\omega})S_y(\omega)\Psi(e^{\imath\omega})^*\,d\omega \tag{3.3}$$

where $S_y(\omega)$ is the spectral density of Y_t and the conditions of Theorem 1 imply that

$$\underline{C}I \leq S_y(\omega) \leq \overline{C}I, \quad 0 < \underline{C} \leq \overline{C} < \infty\,, \tag{3.4}$$

as in the stationary and invertible *VARMA* case. Here and below, as with C, $\underline{C}$ and $\overline{C}$ denote

finite positive constants, not always the same ones. It follows from (3.3) and (3.4), firstly that

$$\|\Sigma_n - \Sigma\|^2 \leq \left(\overline{C} \sum_{i=n+1}^{\infty} \|\Psi_i\|^2 \right)^2 ,$$

and secondly that

$$\underline{C} \sum_{i=n+1}^{\infty} \|\Psi_i\|^2 \leq \mathrm{tr}\{\Sigma^{-1}(\Sigma_n - \Sigma)\} \leq \overline{C} \sum_{i=n+1}^{\infty} \|\Psi_i\|^2 , \tag{3.5}$$

as can be seen by substituting $\Psi(e^{\imath\omega}) + \{\Psi_n(e^{\imath\omega}) - \Psi(e^{\imath\omega})\}$ for $\Psi_n(e^{\imath\omega})$ in (3.3) and employing Parseval's relation and Baxter's inequality, observing that

$$\int_{-\pi}^{\pi} \Psi(e^{\imath\omega}) S_y(\omega) \{\Psi_n(e^{\imath\omega}) - \Psi(e^{\imath\omega})\}^* \, d\omega = 0$$

by virtue of the population ensemble Yule-Walker relations $\Gamma(h) - \sum_{i=1}^{\infty} \Psi_i \Gamma(h-i) = \delta_{0h}\Sigma$, $h = 0, 1, \ldots,$.

Putting the orders of magnitude of $\|\hat{\Sigma}_n - \Sigma_n\|$ and $\|\Sigma_n - \Sigma\|$ together, applying the triangular inequality $\|\hat{\Sigma}_n - \Sigma\| \leq \|\hat{\Sigma}_n - \Sigma_n\| + \|\Sigma_n - \Sigma\|$, and inserting these into (3.2) yields the following result.

Proposition 1 *Let Y_t be as in Theorem 1. If $(T \log T)^{1/2}/P_T \to 0$ then*

$$\rho_T(n) = \log \det \Sigma + tr\{\Sigma^{-1}(\Sigma_n - \Sigma)\} + o(n^{-1}) + \frac{nk^2 P_T}{T}\{1 + o(1)\} . \tag{3.6}$$

uniformly in $n \leq N_T$.

Equation (3.6) indicates that $n_T = \arg\min_{0 \leq n \leq N_T} \rho_T(n)$ is effectively equivalent to ν_T where ν_T minimizes $\bar{\rho}_T(n) = \mathrm{tr}\{\Sigma^{-1}(\Sigma_n - \Sigma)\} + nk^2 P_T/T$, that is, $n_T/\nu_T \to 1$ as $T \to \infty$. If $\sum_{i=1}^{\infty} i^{1/2}\|\Psi_i\| < \infty$ then from (3.5) we have $\underline{C} n^{-1} \leq \mathrm{tr}\{\Sigma^{-1}(\Sigma_n - \Sigma)\} \leq \overline{C} n^{-1}$ and by the squeezing principle $\nu_T = \lfloor (CT/k^2 P_T)^{1/2} \rfloor$ for some C. The upshot of this is that $\|\Psi_{n_T}(z) - \Psi(z)\| = O\{(P_T/T)^{1/4}\}$. Parallel manipulations in the *VARMA* case lead to the conclusion that ν_T solves an equation of the form $C 2 \log|\zeta| \exp(2n \log|\zeta|) + k^2 P_T/T = 0$ and

$$\nu_T = \frac{\log T}{-2 \log|\zeta|}\{1 + o(1)\} ,$$

so that in the *VARMA* case the approximation error

$$\|\Psi_{n_T}(z) - \Psi(z)\| \leq C \exp(\nu_T \log|\zeta|) \leq C T^{-1/2} . \tag{3.7}$$

In real world applications it is common practice to choose n using a model selection criterion such as *AIC*, *HQ* or *BIC*. The penalty terms of these criteria are such that $P_T/(T \log T)^{1/2} < 1$, indicating that n_T determined from *AIC*, *HQ* and *BIC* can be expected to exceed $\nu_T =$

$\arg\min_{0\le n\le N_T}\bar{\rho}_T(n)$, since n_T is a nonincreasing function of P_T (Lütkepohl, 2007, Lemma 4.1). Hence n_T will be of order $-\log T/2\log|\zeta|$ asymptotically in the *VARMA* case. Simulation results suggest, however, that $\nu_T^\zeta = \lfloor -\log T/2\log|\zeta|\rfloor$ overestimates n_T, particularly when ζ is near the unit circle, and that ν_T yields a much better guide to the order of the *VAR(n)* model likely to be chosen in practice. For the RBC model employed below, for example, a simulated realization of $T = 20,000$ observations produced the values 24, 12 and 3 for n_T when using *AIC*, *HQ* and *BIC*, respectively. These are almost identical to the values of 25, 12 and 4 given by ν_T, compared to $\nu_T^\zeta = 110$, see Table 1 below.

4 Structural VAR and Error Decomposition

An examination of the impact of economic shocks on aggregate macroeconomic variables is an important goal in the context of DSGE and RBC modeling, the empirical evidence for which is often obtained by estimating a *SVAR*. The responses of the macroeconomic variables to a structural disturbance are generated by calculating the *IRF* relating $\{Y_t\}$ to the orthogonal structural shocks vector ε_t. To compute the *IRF* in terms of ε_t an identifying matrix Λ such that $u_t = \Lambda\,\varepsilon_t$ is required. Given Λ we can then invert $\Psi(z)$ and express $\{Y_t\}$ as an infinite sum of past structural shocks,

$$Y_t = \sum_{i=0}^{\infty}\Upsilon_i\,\Lambda\,\varepsilon_{t-i}\,, \tag{4.1}$$

where $\Upsilon(z) = I + \sum_{i=1}^{\infty}\Upsilon_i z^i = \Psi(z)^{-1}$ and, by assumption, the variance-covariance matrix of the structural shocks equals the identity. The response of the r'th variable in period $t+j$ to the s'th structural shock at time t is the (r,s) element in the matrix $\Upsilon_j\Lambda$, $j = 0,1,2,\cdots$.

If a *VAR(n)* model is fitted to data as described above then the natural estimate of the transformation matrix Λ is given by $\hat{\Lambda}_n$ where $\hat{\Lambda}_n\hat{\Lambda}_n' = \hat{\Sigma}_n$ and $\hat{\Lambda}_n$ satisfies the same constraints as does Λ. The error incurred from estimating the *IRF* using the *SVAR*(n) is then the difference between $\hat{\Upsilon}_n(z)\hat{\Lambda}_n$, where $\hat{\Upsilon}_n(z) = \hat{\Psi}_n(z)^{-1}$, and $\Upsilon(z)\,\Lambda$, which can be decomposed into the sum of two parts as

$$\hat{\Upsilon}_n(z)\hat{\Lambda}_n - \Upsilon(z)\,\Lambda = \{\hat{\Upsilon}_n(z) - \Upsilon(z)\}\Lambda + \hat{\Upsilon}_n(z)\{\hat{\Lambda}_n - \Lambda\}\,. \tag{4.2}$$

Following Ravenna (2007) we will refer to the first term on the right hand side of (4.2) as the "truncation bias" and to the second as the "identification bias" .

Decomposing $\{\hat{\Upsilon}_n(z) - \Upsilon(z)\}\Lambda$ into $\{\hat{\Upsilon}_n(z) - \Upsilon_n(z)\}\Lambda + \{\Upsilon_n(z) - \Upsilon(z)\}\Lambda$ where $\Upsilon_n(z) = \Psi_n(z)^{-1}$ we find that the "truncation bias" can be broken down into an estimation error component and an approximation error component, and applying the triangular inequality once more gives us

$$\|\{\hat{\Upsilon}_n(z) - \Upsilon(z)\}\Lambda\| \le \|\Lambda\|(\|\hat{\Upsilon}_n(z) - \Upsilon_n(z)\| + \|\Upsilon_n(z) - \Upsilon(z)\|)\,,$$

indicating that the size of the "truncation bias" will be governed by that of the two error components. From Lemma A.1 of Poskitt (2000) we can conclude that

$$\|\hat{\Upsilon}_n(z) - \Upsilon_n(z)\| = O\left(\frac{\|\hat{\Psi}_n(z) - \Psi_n(z)\|}{\|\hat{\Psi}_n(z)\| \cdot \|\Psi_n(z)\|}\right) \tag{4.3}$$

and

$$\|\Upsilon_n(z) - \Upsilon(z)\| = O\left(\frac{\|\Psi_n(z) - \Psi(z)\|}{\|\Psi_n(z)\| \cdot \|\Psi(z)\|}\right). \tag{4.4}$$

The expressions in (4.3) and (4.4) imply that the order of magnitude of the estimation error and approximation error components of the "truncation bias" will be the same as those of the underlying $VAR(n)$ model, save that the constants necessary to accurately prescribe the size of the errors have not been identified.

The "identification bias" can be similarly decomposed into the sum of $\hat{\Upsilon}_n(z)\{\hat{\Lambda}_n - \Lambda_n\}$ and $\hat{\Upsilon}_n(z)\{\Lambda_n - \Lambda\}$ where $\Sigma_n = \Lambda_n\Lambda_n'$ to give, in an obvious manner,

$$\|\hat{\Upsilon}_n(z)\{\hat{\Lambda}_n - \Lambda\}\| \leq \|\hat{\Upsilon}_n(z)\|(\|\hat{\Lambda}_n - \Lambda_n\| + \|\Lambda_n - \Lambda\|).$$

Once again we find that the source of the "identification bias" lies in two parts, an estimation error component and an approximation error component.

5 An RBC Illustration

To evaluate the consequences of using different techniques a common methodology adopted in the literature is to simulate data from a theoretical model by parameterizing the model, solving the log-linearized system around the steady state to get the state-space form solution, and to then generate observations from the theoretical solution – in most cases with short data paths similar to those commonly found in empirical study, i.e. around 200 observations for quarterly data. Performance is then evaluated by comparing various quantities of interest, such as IRFs, computed from the data with their theoretical counterparts. We will apply this method here to the RBC model of Hansen (1985).

There are two exogenous structural shocks in the model: a non-stationary technology shock Z_t, and a stationary labor supply shock D_t. The per-period preference of the representative household is given by the quasilinear form: $\ln C_t + AD_t(1 - N_t)$. The social planner chooses the sequences of consumption goods C_t, capital stock K_t, and labor supply N_t to maximize the expected value of discounted lifetime utility

$$\mathbb{E}_0 \sum_{t=1}^{\infty} \beta^t [\ln C_t + AD_t(1 - N_t)], \tag{5.1}$$

subject to the capital accumulation law and a Cobb-Douglas production technological constrain-

t:

$$K_t = Y_t - C_t + (1-\delta)K_{t-1}, \tag{5.2}$$

$$Y_t = K_{t-1}^{\alpha}(Z_t N_t)^{1-\alpha}. \tag{5.3}$$

Denote the gross interest rate R_t as

$$R_t = \alpha \frac{Y_t}{K_{t-1}} + (1-\delta). \tag{5.4}$$

The first order conditions for the maximization problem are

$$AD_t = (1-\alpha)\frac{Y_t}{C_t N_t} \qquad \text{and} \qquad \beta\, \mathbb{E}_t \left\{ \frac{C_t}{C_{t+1}} R_{t+1} \right\} = 1. \tag{5.5}$$

The labor-augmenting technology level Z_t and the labor supply shifter D_t follow exogenous stochastic processes:

$$\ln Z_t = \ln Z_{t-1} + \mu_z + \eta_{z_t}; \tag{5.6}$$

$$\ln D_t = (1-\rho_d)\ln \bar{D} + \rho_d \ln D_{t-1} + \eta_{d_t}, \tag{5.7}$$

where μ_z is the drift term for the random-walk process $\{Z_t\}$, $\bar{D}$ denotes the long-run mean of D_t, $\eta_{at} \sim \mathbf{N}(0, \sigma_a^2)$, $a = z, d$, and $\rho_d \neq 0$. The economy dynamic equilibrium is summarized in equations (5.2)–(5.7).

A technology shock has a permanent effect on the level of Z_t, and hence on C_t, K_t, and Y_t. We have therefore defined the model in terms of $\{N_t, R_t, D_t, \hat{C}_t = C_t/Z_t, \hat{Y}_t = Y_t/Z_t, \hat{K}_t = K_t/Z_t, \hat{Z}_t = Z_t/Z_{t-1}\}_{t=1}^{\infty}$ so as to make all the series stationary, and the system is solved by log-linearizing (5.2)–(5.7) around the steady state. For any variable X_t, define its log-deviation from steady state by the lower case letter $x_t = \ln(X_t/\bar{X})$, so that $X_t = \bar{X}e^{x_t} \approx \bar{X}(1+x_t)$. Following Blanchard and Quah (1989) hours worked n_t and output growth $\Delta y_t = \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t$ are taken as the observable variables, so that in our previous notation $Y_t := (n_t \quad \Delta y_t)'$.

5.1 Population Ensemble Parameterization and Evaluations

The model is parameterized according to Erceg, Guerrieri, and Gust (2005) and Ravenna (2007). First the mean of the technology shock is set to $\mu_z = 0.0037$ and the variance $\sigma_z = 0.0148$, hence $\overline{\hat{Z}} = e^{0.0037}$. The steady state level of the labor supply shock is normalized to 1, since it does not affect the model's log-linear dynamics. The variance $\sigma_d = 0.009$. The first order autocorrelation coefficient $\rho_d = 0.80$, which indicates relatively strong persistence for the labor supply shock. The total capital share α is set to 0.35. The quarterly depreciation rate for installed capital δ is assumed to equal 2% and $\beta = 1.03^{-0.25}$. The total labor endowment is normalizing to be one, and the steady state level of labor $\bar{N}$ is set to 1/3.

These values lead to the following *VARMA* representation of the data generating process

$$Y_t = \begin{pmatrix} 0.9413 & 1.0446 \\ 0.00060 & 0.8045 \end{pmatrix} Y_{t-1} + u_t + \begin{pmatrix} -0.2498 & -0.9173 \\ -0.1924 & -0.7065 \end{pmatrix} u_{t-1} \tag{5.8}$$

where the reduced form disturbance $u_t = B\eta_t$ where

$$B = \begin{pmatrix} 0.4821 & -2.4030 \\ 0.9634 & -1.5619 \end{pmatrix}$$

and the vector structural shock $\eta_t = (\eta_{z_t} \quad \eta_{d_t})'$. The variance-covariance matrix of the reduced form disturbance

$$\Sigma = BDB' = \begin{pmatrix} 0.51863008 & 0.40575031 \\ 0.40575031 & 0.40089860 \end{pmatrix} 10^{-3}, \tag{5.9}$$

where $D = \text{diag}\{\sigma_z^2, \sigma_d^2\}$.

From the stochastic structure as specified in (5.8) and (5.9) the population ensemble parameters can now be evaluated. First the autocovariances $\Gamma(h) = \Gamma(-h)'$, $h = 0, 1, 2, \ldots$, can be evaluated using the numerical procedures outlined in the Appendix. Thus, the initial autocovariance matrices

$$\Gamma(0) = \begin{pmatrix} 0.0455 & 0.0026 \\ 0.0026 & 0.0010 \end{pmatrix} \qquad \text{and} \qquad \Gamma(1) = \begin{pmatrix} 0.0447 & 0.0028 \\ 0.0015 & 0.0003 \end{pmatrix}$$

are easily computed, with the higher order autocovariance matrices $\Gamma(h)$, $h > 1$, being readily evaluated recursively via $\Gamma(h) = \Phi_1\Gamma(h-1)$ where, of course,

$$\Phi_1 = \begin{pmatrix} 0.9413 & 1.0446 \\ 0.00060 & 0.8045 \end{pmatrix}.$$

For any $n = 0, 1, 2, \cdots$, the Levinson–Whittle algorithm is then used to calculate $\Psi_{n1}, \cdots, \Psi_{nn}$ and Σ_n, the parameters of the population ensemble MMSE $VAR(n)$ approximation.

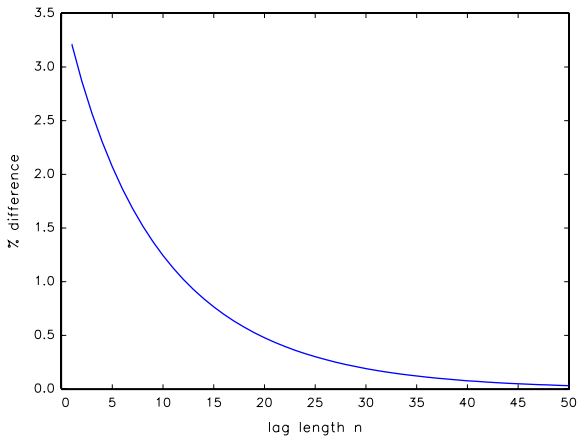
We use the identification scheme of Blanchard and Quah (1989) to construct the transformation matrix from the reduced form residuals to the orthogonal structural shocks necessary to formulate a *SVAR*. In the RBC model η_{d_t} has no long-run effect on either employment n_t or total production y_t, and the technology shock η_{z_t} has no long-run effect on employment, but has a long-run effect on total production. The *IRF* of y_t to an orthonormalized labor supply shock is derived from the coefficients in powers of z in $[\Upsilon(z)\Lambda]_{12}$ where $\Upsilon(z) = \Psi(z)^{-1} = \Phi(z)^{-1}\Theta(z)$, and imposition of the identification constraint that the long-run effect of a labor supply shock on Δy_t is zero implies that $[\Upsilon(1)\Lambda]_{12} = 0$. The covariance matrix of the reduced form disturbance equals $\Sigma = BDB'$ and the Cholesky factorization of $\Upsilon(1)\Sigma\Upsilon(1)' = \Upsilon(1)BDB'\Upsilon(1)'$ yields a lower-triangular matrix H such that $HH' = \Upsilon(1)\Sigma\Upsilon(1)'$, implying that the matrix $\Lambda = \Upsilon(1)^{-1}H$ is such that the $(1,2)$ element of $\Upsilon(1)\Lambda$ is zero, as required by the long run

identification assumption, and $\Lambda\Lambda' = BDB'$, as required for orthonormalization. For a $VAR(n)$ approximation, the same procedure is used to identify Λ_n, namely $\Lambda_n = \Upsilon_n(1)^{-1}H_n$, where H_n is the lower-triangular Cholesky factor such that $H_nH_n' = \Upsilon_n(1)\Sigma_n\Upsilon_n(1)'$. In the RBC model a positive technology shock has a positive effect on total output, and the sign of the impulse responses is identified by matching the direction of the long-run impact of a technology shock on output.

5.2 VAR(n) Model Convergence

The covariance matrix Σ_n of a $VAR(n)$ approximation is monotonically nonincreasing in n and approaches Σ, the innovations covariance, as $n \to \infty$. In Figure 1 we plot $d_n = \|\Sigma^{-1}(\Sigma_n - \Sigma)\|$ as a function of the lag length n. We can think of d_n as a goodness-of-fit measure that measures the percentage difference between Σ_n and Σ.

Figure 1: Plot of percentage difference $d_n = \|\Sigma^{-1}(\Sigma_n - \Sigma)\|$ as function of n.

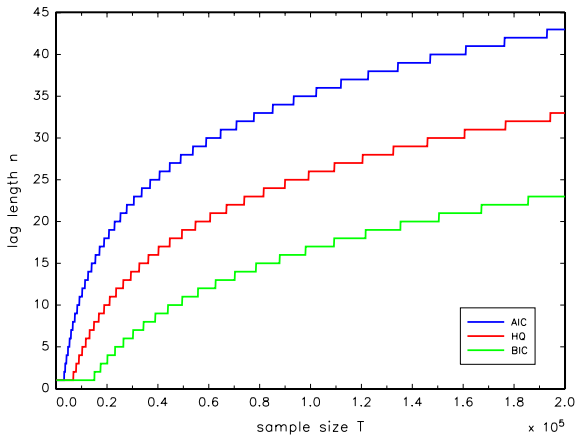


From Figure 1 it is evident that $d_n < 2\%$ for n as small as 5, and that Σ_n is within 1% of Σ by the time $n = 13$. This rapid decline in d_n is reflected in the values of ν_T seen in Figure 2, which graphs ν_T for $P_T = 2, 2\log\log T$ and $\log T$, the penalty terms associated with AIC, HQ and BIC respectively, for T between 200 and 200,000.

The value of ν_T is very small for all but the very largest values of T. Even for AIC, the most profligate criterion, T needs to be at least 10,000 before ν_T will exceed 10, $\nu_T = 20$ when $T = 23,000$, and to produce a value of ν_T greater than 45 requires that T exceed 200,000.

That these figures have practical relevance is seen in Table 1. This table presents the values of n_T determined by AIC, HQ and BIC from a realization of the process in (5.8) generated using i.i.d. zero mean Gaussian reduced form disturbances with covariance as in (5.9), and compares these with ν_T and the basic asymptotic approximation ν_T^ζ. For each criterion we find that $\nu_T \leq n_T < \nu_T^\zeta$ on all bar one occasion, and n_T and ν_T are reasonably close and considerably less than ν_T^ζ, indicating that $n_T/\nu_T \to 1$ as $T \to \infty$ much more rapidly than does n_T/ν_T^ζ. The values displayed in Figure 2 and listed in Table 1 lend strong theoretical support

Figure 2: Asymptotic orders ν_T of VAR approximations generated by AIC, HQ and BIC.



to the observations of Chari, Kehoe, and McGrattan (2007) and Kapetanios, Pagan, and Scott (2007) concerning the orders of $VAR(n)$ approximations found in practice in the context of RBC and DSGE modeling.

Table 1: Orders of $VAR(n)$ approximations

T	AIC		HQ		BIC		
	n_T	ν_T	n_T	ν_T	n_T	ν_T	ν_T^{ζ}
200	1	1	1	1	1	1	59
400	1	1	1	1	1	1	66
800	1	1	1	1	1	1	74
1,600	8	1	1	1	1	1	82
3,200	8	2	1	1	1	1	89
5,000	13	5	1	1	1	1	94
8,000	17	9	2	3	1	1	100
12,000	24	13	6	6	1	1	104
20,000	24	18	12	10	3	3	110

To illustrate the impact of using different order $VAR(n)$ approximations we present in Table 2 the values of $T^{1/2}\|\hat{\Psi}_n(z) - \Psi_n(z)\|/n(\log T)^{1/2}$ and $T^{1/2}\|\Psi_n(z) - \Psi(z)\|$, the normalized estimation and approximation errors, obtained when $n = \nu_T$ is calculated using $P_T = 2, 2\log\log T$ and $\log T$, corresponding to AIC, HQ and BIC respectively. To allow for sampling fluctuation the estimation error is averaged across $R = 1000$ replications. Given that the sequences comprised by $T^{1/2}\|\hat{\Psi}_n(z) - \Psi_n(z)\|/n(\log T)^{1/2}$ and $T^{1/2}\|\Psi_n(z) - \Psi(z)\|$ have to have limit points constituting precisely an interval $[0, C]$ if the order of magnitude is to be deemed to hold, a perhaps surprising feature of Table 2 is that the normalized approximation error increases steadily with T, over the range of T considered, and is always larger than the normalized estimation error, which first increases and then decreases and stabilizes. Such outcomes indicate that the bound in Theorem 1 starts to bind more quickly as $T \to \infty$ than does (3.7) and they suggest

that an extremely large value of T indeed may be required before (3.7) becomes relevant.

Table 2: $VAR(n)$ normalized estimation error and approximation error, $Est. = T^{1/2}\|\hat{\Psi}_{\nu_T}(z) - \Psi_{\nu_T}(z)\|/\nu_T(\log T)^{1/2}$ and $Approx. = T^{1/2}\|\Psi_{\nu_T}(z) - \Psi(z)\|$.

T	AIC		HQ		BIC	
	Est.	*Approx.*	*Est.*	*Approx.*	*Est.*	*Approx.*
200	7.0131	7.9918	7.0131	7.9918	7.0131	7.9918
400	9.3522	11.3021	9.3522	11.3021	9.3522	11.3021
800	12.5171	15.9836	12.5171	15.9836	12.5171	15.9836
1,600	16.8520	22.6042	16.8520	22.6042	16.8520	22.6042
3,200	9.8112	30.2669	22.7719	31.9672	22.7719	31.9672
5,000	5.0384	33.0199	27.7075	39.9590	27.7075	39.9590
8,000	3.6052	34.9150	14.6720	47.8561	34.1407	50.5445
12,000	3.0817	35.8042	6.2435	48.9030	40.8945	61.9041
20,000	2.8706	37.0607	4.8990	52.8021	14.9985	72.2974

This phenomenon is partly a function of the fact that, for the RBC model under consideration here, the smallest root of $\det\Theta(z)$, 1.0459, is near the unit circle. This implies that the theoretical autoregressive coefficients in (1.2) decline slowly. Nevertheless, the minimizing values $\Psi_{n1}, \cdots, \Psi_{nn}$ in the population ensemble MMSE $VAR(n)$ approximation are such as to make Σ_n close to Σ for small values of n. Now, the convergence rate of the estimators $\hat{\Psi}_{n1}, \cdots, \hat{\Psi}_{nn}$ and $\hat{\Sigma}_n$ depends on convergence characteristics of the sample autocovariances and is in essence only a function of n and T. Taken together these two properties lead to both the order and the estimation error of the fitted $VAR(n)$ approximation being small at moderate to large sample sizes. But for small to moderate values of n it is difficult for a $VAR(n)$ approximation to capture the slow decay in the $VAR(\infty)$ coefficients and $\sum_{i=1}^{n}\|\Psi_{ni} - \Psi_i\|^2$ and $\sum_{i=n+1}^{\infty}\|\Psi_i\|^2$ remain large, and consequently the approximation error does not decrease as fast as the asymptotic theory dictates as T increases. Thus in practice the approximation error approaches its asymptotic position far more slowly than does the estimation error, their theoretical orders of magnitude notwithstanding.

5.3 Impulse Response Analysis and SVARs

Table 3 is the counterpart to Table 2 and for the corresponding $VAR(n)$ models reports the values of the normalized estimation and approximation errors $T^{1/2}\|\{\hat{\Upsilon}_n(z) - \Upsilon_n(z)\}\Lambda\|/n(\log T)^{1/2}$ and $T^{1/2}\|\{\Upsilon_n(z) - \Upsilon(z)\}\Lambda\|$. These two values measure the size of the estimation and approximation error components of the *IRF* "truncation bias" . As previously, the normalized approximation error increases with increasing T and the normalized estimation error increases and then eventually decreases, the behaviour of the estimation error and approximation error components of the "truncation bias" reflecting that of these two components in the *VAR* coefficients.

To illustrate the source of the approximation error we present in Figure 3 the *IRF*s of Δy_t and n_t to a orthonormalized technology shock η_{z_t}. The figure plots the *IRF* generated by

Table 3: *IRF* "truncation bias" normalized estimation and approximation error, $Est. = T^{1/2}\|\hat{\Upsilon}_{\nu_T}(z) - \Upsilon_{\nu_T}(z)\Lambda\|/\nu_T(\log T)^{1/2}$ and $Approx. = T^{1/2}\|\{\Upsilon_{\nu_T}(z) - \Upsilon(z)\}\Lambda\|$.

T	AIC		HQ		BIC	
	Est.	*Approx.*	*Est.*	*Approx.*	*Est.*	*Approx.*
200	27.8860	20.1053	27.8860	20.1053	27.8860	20.1053
400	37.0346	28.4333	37.0346	28.4333	37.0346	28.4333
800	49.6265	40.2107	49.6265	40.2107	49.6265	40.2107
1,600	66.7806	56.8665	66.7806	56.8665	66.7806	56.8665
3,200	48.7842	77.4968	90.2831	80.4214	90.2831	80.4214
5,000	24.5628	85.2135	109.8573	100.5267	109.8573	100.5267
8,000	17.1695	88.8078	73.0751	122.5331	135.2409	127.1574
12,000	14.3850	88.6897	30.3945	125.9882	162.0459	155.7353
20,000	13.1392	88.4937	23.3340	133.5055	74.4045	186.0921

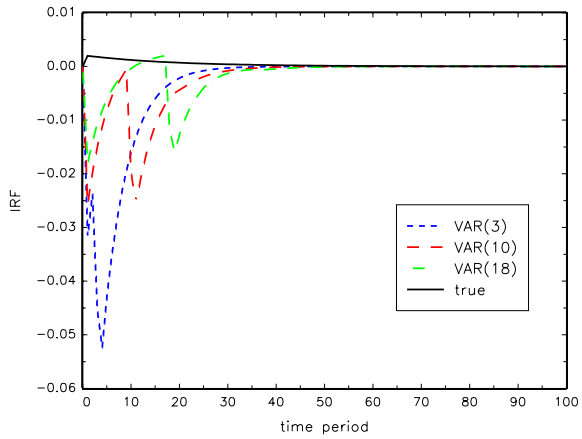
three theoretical $VAR(n)$ approximations and compares them to the true *IRF*. The values of n used are drawn from those seen in Table 1. For each lag length employed the profile of the impulse responses from the *VAR(n)* approximation, $\Upsilon_n(z)\Lambda_n$, clearly differ considerably from those from $\Upsilon(z)\Lambda$, the true impulse responses, and these differences accumulate to produce the approximation error.

In practice estimated impulse responses produced by fitted $VAR(n)$ models will be centered around $\Upsilon_n(z)\Lambda_n$, but it is well known that the standard errors of estimated impulse responses can be disappointingly large (Lütkepohl, 1990), raising the possibility that confidence bands set around $\hat{\Upsilon}_n(z)\hat{\Lambda}_n$ may yet cover the true *IRF*. That the probability of the latter event may not be insignificant follows from observing that although the estimation error approaches its asymptotic position faster than does the approximation error, it is the estimation error that is ultimately the more dominant component of both the "truncation bias" and the "identification bias". The latter feature is readily verified upon noting that the normalized errors reported in Table 3 imply that the relative error $\|\{\hat{\Upsilon}_n(z) - \Upsilon_n(z)\}\Lambda\|/\|\{\Upsilon_n(z) - \Upsilon(z)\}\Lambda\|$ exceeds 3 and can be as large as 8 when $T = 20,000$. Similar calculations not detailed here – but see the following paragraph – show that $\|\hat{\Upsilon}_n(z)\{\hat{\Lambda}_n - \Lambda_n\}\|/\|\hat{\Upsilon}_n(z)\{\Lambda_n - \Lambda\}\|$ exceeds 2 and can reach 12 when $T = 20,000$.

In empirical applications researchers are often interested in the long run effects of structural shocks on the economy. In Table 4 we therefore present the "identification bias" component of the long run variance estimate broken down into the normalized estimation and approximation errors. The patterns seen previously, the gradual increase in the normalized approximation error and the rise and fall in the normalized estimation error, repeat themselves here. The numerical values seen in Table 4 show, however, that the constants necessary to prescribe the size of the errors are much larger than before, indicating that the estimation and approximation error components in the *IRF* and the covariance estimates do not cancel but are compounded in the construction of the *SVAR*.

Figure 3: *IRF* from theoretical $VAR(n)$ approximations.

(a) *IRF* of Δy_t to technology shock.



(b) *IRF* of n_t to technology shock.

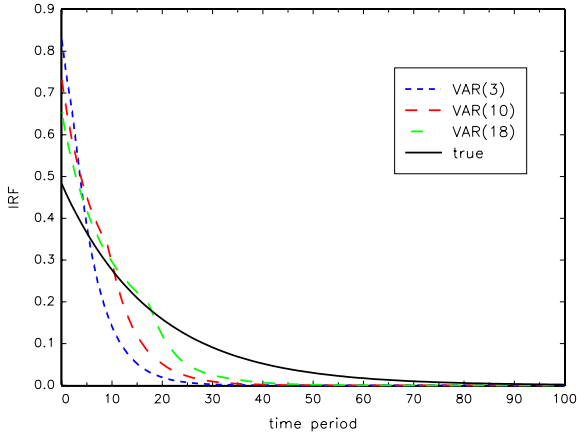


Table 4: Normalized estimation and approximation error of long run variance "identification bias" , $Est. = T^{1/2}\|\hat{\Upsilon}_{\nu_T}(1)\{\hat{\Lambda}_{\nu_T} - \Lambda_{\nu_T}\}\|/\nu_T(\log T)^{1/2}$ and $Approx. = T^{1/2}\|\hat{\Upsilon}_{\nu_T}(1)\{\Lambda_{\nu_T} - \Lambda\}\|$.

T	AIC		HQ		BIC	
	Est.	*Approx.*	*Est.*	*Approx.*	*Est.*	*Approx.*
200	41.1112	39.5137	41.1112	39.5137	41.1112	39.5137
400	54.1129	55.4701	54.1129	55.4701	54.1129	55.4701
800	72.4482	78.2990	72.4482	78.2990	72.4482	78.2990
1,600	97.2356	110.4659	97.2356	110.4659	97.2356	110.4659
3,200	75.6898	151.1469	131.0681	155.6892	131.0681	155.6892
5,000	49.9495	170.4261	159.5249	194.6706	159.5249	194.6706
8,000	43.6972	184.0925	113.4554	239.4554	196.6725	246.7042
12,000	42.1736	190.0239	66.2557	254.8612	235.5969	301.9944
20,000	43.5464	196.0537	61.9112	279.1760	128.2178	366.4671

6 Conclusion

In this article we have investigated the consequences of fitting $VAR(n)$ models to time series that in truth come from a $VAR(\infty)$ data generating mechanism. We have shown that overall error can be broken down into two basic components, an estimation error that stems from the difference between the parameter estimates and their population ensemble MMSE $VAR(n)$ approximation counterparts, and an approximation error that stems from the difference between the MMSE $VAR(n)$ approximation and the true $VAR(\infty)$. This dichotomy of error permeates through to other performance indicators previously employed in the literature, such as the "truncation bias" and "identification bias" of $SVAR$s, and the two sources of error cannot be ignored.

Our theoretical analysis indicates that the magnitude of the estimation error exceeds that of the approximation error, but experimental results based upon a prototypical RBC model indicate that in practice the approximation error approaches its asymptotic position far more slowly than does the estimation error, their relative orders of magnitude notwithstanding. Treating the results obtained in our experiments as a counter example implies that the commonly employed justification for using a $VAR(n)$ model – that a $VAR(n)$ will approximate a $VAR(\infty)$ process arbitrarily closely in the limit, that is, as $n \to \infty$ as $T \to \infty$ – may not be applicable unless T is enormous. Moreover, whereas in practice the consistency property of the estimators is manifest when T is reasonably large, the approximation error does not disappear as fast as the asymptotic theory dictates, suggesting that with sample sizes and lag lengths like those commonly employed in practice $VAR(n)$ models are likely to exhibit serious errors of both types and behave poorly. That inferences based on $VAR(n)$ models can be untrustworthy is a finding in close accord with the conclusions of others, but that the poor performance cannot be attributed to a single source of error is a feature that does not appear to have been appreciated in the previous literature.

In any empirical application it is not possible to suppress estimation error, so given the presence of both estimation error and approximation error in VARs it is natural to contemplate modifying the class of models employed in practice so as to shut down the latter error. This can be achieved by estimating the underlying parameters of a completely specified and accepted model using likelihood based techniques, as outlined in Komunjer and Ng (2011). Alternatively, if economists wish to identifying economic shocks and their impulse response functions from an incompletely specified reduced form that could be derived from the structural model of interest, and which can account for the stylized facts observed with macroeconomic variables, they can proceed by fitting $VARMA$ models using the methodology described in Poskitt (2011). Inferential methods associated with VAR modeling are currently more familiar and more easily applied using standard software than are those associated with these alternative approaches, needless to say, but given growing evidence of the type presented in this paper on the unreliability of VAR models an argument for the continued use of VARs based on familiarity and ease of implementation seems unwarranted.

Appendix: Calculation of Population Ensemble Parameters

For any *VARMA* process defined as in equation (1.1), henceforth denoted by *VARMA*(Φ, Θ, Σ), the autocovariance generating function (AGF) is given by $\Gamma(z) = \Upsilon(z)\Sigma\Upsilon(z^{-1})'$. Contrary to the statement made in Komunjer and Ng (2011, p.2001), the coefficients in $\Gamma(z) = \sum_{h=-\infty}^{\infty}\Gamma(h)z^h$, namely

$$\Gamma(h) = \Gamma(-h)' = \sum_{i=0}^{\infty}\Upsilon_{h+i}\Sigma\Upsilon_i', \quad h = 0, 1, 2, \dots,$$

can be evaluated without having to resort to truncation and using partial sum approximations to the series expansions. This is achieved by noting that $\Phi(z)\Gamma(z) = \Theta(z)\Sigma\Upsilon(z^{-1})'$ and equating coefficients in the latter expression yields the equations

$$\Gamma(h) - \sum_{i=1}^{p}\Phi_i\Gamma(h-i) = \begin{cases} \sum_{h\le j\le q}\Theta_j\Sigma\,\Upsilon_{j-h}' & \text{for } 0 \le h \le q \\ 0 & \text{for all } h > q\,. \end{cases} \tag{A.1}$$

Let $s = \max\{p, q\}$. The first $s+1$ equations in (A.1) can be solved for $\Gamma(0), \cdots, \Gamma(s)$ and the remaining equations then give $\Gamma(h) = \Phi_1\Gamma(h-1) + \cdots + \Phi_p\Gamma(h-p)$ recursively for any $h > s$.

To initiate the calculations first set $\Phi_i = 0$, $i = p+1, \dots, s$, $p < s$, or $\Theta_i = 0$, $i = q+1, \dots, s$, $q < s$, and set the $k(s+1) \times k$ matrix $\boldsymbol{\Upsilon}_s = (I_k, \Upsilon_1', \dots, \Upsilon_s')'$. The *IRF* coefficients in $\boldsymbol{\Upsilon}_s$ can be computed from the recursion $\Upsilon_j = \sum_{i=1}^{j}\Phi_i\Upsilon_{j-i} + \Theta_j$, $j = 1, \dots, s$, initiated at $\Upsilon_0 = I_k$, or equivalently using basic linear algebra from the equation

$$\boldsymbol{\Upsilon}_s = \left(I_{k(s+1)} - \boldsymbol{\Phi}_T\right)^{-1}\boldsymbol{\Theta} \tag{A.2}$$

where $\boldsymbol{\Phi}_T$ is the $k(s+1) \times k(s+1)$ matrix given by

$$\boldsymbol{\Phi}_T = \begin{pmatrix} \mathbf{0}_{k\times ks} & 0_{k\times k} \\ \mathbf{T}_\Phi & 0_{ks\times k} \end{pmatrix} \quad \text{where} \quad \mathbf{T}_\Phi = \begin{pmatrix} \Phi_1 & 0 & \cdots & 0 \\ \Phi_2 & \Phi_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_s & \Phi_{s-1} & \cdots & \Phi_1 \end{pmatrix},$$

and $\boldsymbol{\Theta}' = (I_k, \Theta_1', \dots, \Theta_s')'$, see Mittnik (1987).

Now set

$$\boldsymbol{\Gamma}^+ = \begin{pmatrix} \Gamma(0) \\ \Gamma(1) \\ \vdots \\ \Gamma(s) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}^- = \begin{pmatrix} \Gamma(0) \\ \Gamma(-1) \\ \vdots \\ \Gamma(-s) \end{pmatrix}$$

and let

$$\mathbf{\Phi}_H = \begin{pmatrix} \mathbf{0}_{ks\times k} & \mathbf{H}_\Phi \\ \mathbf{0}_{k\times k} & \mathbf{0}_{k\times ks} \end{pmatrix} \quad \text{where} \quad \mathbf{H}_\Phi = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{s-1} & \Phi_s \\ \Phi_2 & \Phi_3 & \cdots & \Phi_s & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Phi_{s-1} & \Phi_s & \cdots & 0 & 0 \\ \Phi_s & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Then it can be shown (Mittnik, 1990, 1993) that $\mathbf{\Gamma}^+$ and $\mathbf{\Gamma}^-$ are related by the formula

$$\mathbf{\Gamma}^+ = \mathbf{\Phi}_T\,\mathbf{\Gamma}^+ + \mathbf{\Phi}_H\mathbf{\Gamma}^- + \mathbf{\Theta}_H\,(I_{s+1}\otimes\Sigma)\,\mathbf{Y}_s \tag{A.3}$$

where

$$\mathbf{\Theta}_H = \begin{pmatrix} I_k & \Theta_1 & \cdots & \Theta_{s-1} & \Theta_s \\ \Theta_1 & \Theta_2 & \cdots & \Theta_s & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Theta_{s-1} & \Theta_s & \cdots & 0 & 0 \\ \Theta_s & 0 & \cdots & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{Y}_s = \begin{pmatrix} I_k \\ \Upsilon_1' \\ \vdots \\ \Upsilon_s' \end{pmatrix}.$$

Transposing and vectorizing equation (A.3) yields

$$\boldsymbol{\gamma} = (\mathbf{\Phi}_T\otimes I_k)\,\boldsymbol{\gamma} + (\mathbf{\Phi}_H\otimes I_k)(I_{s+1}\otimes K)\boldsymbol{\gamma} + \operatorname{vec}(\mathbf{Y}_s'(I_{s+1}\otimes\Sigma)\mathbf{\Theta}_H')\,, \tag{A.4}$$

where $\boldsymbol{\gamma} = \operatorname{vec}([\Gamma(0),\Gamma(-1),\ldots,\Gamma(-s)])$ and K is the $k^2\times k^2$ commutation matrix such that $K\operatorname{vec}(\Gamma(h)) = \operatorname{vec}(\Gamma(h)') = \operatorname{vec}(\Gamma(-h))$. Thus $\boldsymbol{\gamma}$ can be obtained by solving the $k^2(s+1)$th order linear equation system $\mathbf{M}\boldsymbol{\gamma} = \mathbf{m}$ where $\mathbf{M} = I_{k^2(s+1)} - \mathbf{\Phi}_T\otimes I_k - (\mathbf{\Phi}_H\otimes I_k)(I_{s+1}\otimes K)$ and $\mathbf{m} = \operatorname{vec}(\mathbf{Y}_s'(I_{s+1}\otimes\Sigma)\mathbf{\Theta}_H')$.

Given $\Gamma(h) = \Gamma(-h)'$, for $h = 0,1,\ldots,s,s+1,\ldots$, evaluated in the manner just outlined, we can now apply the Levinson–Whittle algorithm (Hannan and Deistler, 1988, p. 218) to calculate for any n the parameters $\Psi_{n1},\ldots,\Psi_{nn}$ and Σ_n of the MMSE *VAR(n)* approximation.

Having once calculated the population ensemble parameters of the MMSE *VAR(n)* approximation, the fitted *VAR(n)* estimation error is readily computed directly as $\|\hat\Psi_n(z) - \Psi_n(z)\| = (\sum_{i=1}^n \|\hat\Psi_{ni} - \Psi_{ni})\|^2)^{1/2}$. The approximation error is evaluated by recognizing that the equations

$$\|\Psi_n(z) - \Psi(z)\| = \|\Psi_n(z) - \Theta(z)^{-1}\Phi(z)\| = \|\Theta(z)^{-1}\{\Theta(z)\Psi_n(z) - \Phi(z)\}\|$$

imply that the norm $\|\Psi_n(z) - \Psi(z)\| = \sqrt{\operatorname{tr}\Gamma(0)}$ where $\Gamma(0)$ now stands for the covariance of a $VARMA(\Theta,\{\Theta\Psi_n - \Phi\},I)$ process. Similarly, some straightforward manipulations show that

$$\|\{\hat\Upsilon_n(z) - \Upsilon_n(z)\}\Lambda\| = \|\frac{\hat\Psi_n(z)^{-1}}{\det\Psi_n(z)}\{\Psi_n(z) - \hat\Psi_n(z)\}\Psi_n^\dagger(z)\Lambda\|\,,$$

where $M^\dagger$ denotes the adjoint of M, and

$$\|\{\Upsilon_n(z) - \Upsilon(z)\}\Lambda\| = \|\frac{\hat{\Psi}_n(z)^{-1}\Theta(z)^{-1}}{\det\Phi(z)}\{\Phi(z) - \Theta(z)\hat{\Psi}_n(z)\}\Phi^\dagger(z)\Theta(z)\Lambda\|\,.$$

Hence the squared norm of the estimation error and the approximation error components of the "truncation bias" equal the trace of the covariance of a $VARMA(\hat{\Psi}_n \det\Psi_n, \{\Psi_n - \hat{\Psi}_n\}\Psi_n^\dagger, \Sigma)$ and a $VARMA(\Theta\hat{\Psi}_n \det\Phi, \{\Phi - \Theta\hat{\Psi}_n\}\Phi^\dagger\Theta, \Sigma)$ process, respectively. Thus the two error components of a *VAR(n)* model can be calculated in closed form using the procedures to determine the AGF of a *VARMA* process given immediately above. By exploiting these relationships in this manner we avoid using numerical methods (based on partial sums of series expansions) that are subject to the very same truncation effects that we are attempting to evaluate and we can compute the estimation and approximation errors with a high degree of numerical accuracy.

Acknowledgement: We are grateful to Farshid Vahid for constructive comments on a preliminary version of this paper.

References

Baxter, G. (1962): "An Asymptotic Result for the Finite Predictor," *¡ath. Scand.*, 10, 137–144.

Blanchard, O. J., and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(4), 655–73.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2007): "Are structural VARs with long-run restrictions useful in developing business cycle theory?," *Federal Reserve Bank of Minneapolis, Research Department Staff Report 364*.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2006): "Assessing Structural VARs," NBER Working Papers 12353, National Bureau of Economic Research, Inc.

DelNegro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARS," *International Economic Review*, 45(2), 643–673.

Erceg, C. J., L. Guerrieri, and C. Gust (2005): "Can Long-Run Restrictions Identify Technology Shocks?," *Journal of the European Economic Association*, 3(6), 1237–1278.

Fernndez-Villaverde, J., J. F. Rubio-Ramandiacute;rez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97(3), 1021–1026.

Hannan, E. J., and M. Deistler (1988): *The Statistical Theory of Linear Systems*. John Wiley.

Hansen, G. D. (1985): "Indivisible labor and the business cycle," *Journal of Monetary Economics*, 16(3), 309–327.

Kapetanios, G., A. Pagan, and A. Scott (2007): "Making a match: Combining theory and evidence in policy-oriented macroeconomic modeling," *Journal of Econometrics*, 136(2), 565–594.

Kascha, C., and K. Mertens (2009): "Business cycle analysis and VARMA models," *Journal of Economic Dynamics and Control*, 33(2), 267 – 282.

Komunjer, I., and S. Ng (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79(6), 1995–2032.

Lippi, M., and L. Reichlin (1994): "VAR analysis, nonfundamental representations, blaschke matrices," *Journal of Econometrics*, 63(1), 307–325.

Lütkepohl, H. (1990): "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *The Review of Economics and Statistics*, 72(1), 116–25.

——— (2007): *The New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

Mittnik, S. (1987): "Non-recursive methods for computing the coefficients of the autoregressive and the moving-average representation of mixed ARMA processes," *Economics Letters*, 23(3), 279 – 284.

——— (1990): "Computation of Theoretical Autocovariance Matrices of Multivariate Autoregressive Moving Average Time Series," *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1), 151–155.

——— (1993): "Computing Theoretical Autocovariances of Multivariate Autoregressive Moving Average Models by Using a Block Levinson Method," *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(2), pp. 435–440.

Poskitt, D. S. (1994): "A Note on Autoregressive Modeling," *Econometric Theory*, 10(5), pp. 884–899.

——— (2000): "Strongly Consistent Determination of Cointegrating Rank Via Canonical Correlations," *Journal of Business and Economic Statistics*, 18, 77–90.

——— (2011): "Vector Autoregressive Moving Average Identification for Macroeconomic Modeling: A New Methodology," Discussion paper, Department of Econometrics & Business Statistics working paper WP12-09, Monash University, Presented at New Developments in Time Series Econometrics Conference, European University Institute, Florence, September 2011.

Ravenna, F. (2007): "Vector autoregressions and reduced form representations of DSGE models," *Journal of Monetary Economics*, 54(7), 2048–2064.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), 1–48.

Smith, A A, J. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8(S), S63–84.