Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 842

October 2005

Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions

Lawrence J. Christiano, Martin Eichenbaum and Robert Vigfusson

Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions*

Lawrence J. Christiano, Martin Eichenbaum, and Robert Vigfusson†

Abstract

We show that the standard procedure for estimating long-run identified vector autoregressions uses a particular estimator of the zero-frequency spectral density matrix of the data. We develop alternatives to the standard procedure and evaluate the properties of these alternative procedures using Monte Carlo experiments in which data are generated from estimated real business cycle models. We focus on the properties of estimated impulse response functions. In our examples, the alternative procedures have better small sample properties than the standard procedure, with smaller bias, smaller mean square error and better coverage rates for estimated confidence intervals.

Keywords: technology shocks, hours worked, frequency domain, spectral density matrix

JEL Codes: E24, E32, O3

1 Introduction

There is a large literature in which researchers impose long-run identifying restrictions on vector autoregressions (VARs) to identify the dynamic effects of shocks to the economy. These restrictions are motivated by the implications that many economic models have for the long-run effects of shocks. We show that the standard procedure for estimating long-run-identified vector autoregressions uses

*The first two authors are grateful for the financial support of grants from the NSF to the National Bureau of Economic Research. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any person associated with the Federal Reserve System.

†Christiano: Northwestern University, email l-christiano@northwestern.edu, Eichenbaum: Northwestern University, email: eich@northwestern.edu, Vigfusson: Board of Governors of the Federal Reserve System, email: robert.j.vigfusson@frb.gov

a particular estimator of the zero-frequency spectral density matrix of the data. This estimator is derived from the estimated coefficients of a finite order VAR. When the actual VAR of the data contains more lags of the variables than the econometrician uses, this estimator of the zero-frequency spectral density matrix can have poor properties. These poor properties stem from the difficulties involved in estimating the sum of the coefficients in a VAR. We develop alternative procedures that combine different zero-frequency spectral density matrix estimators with VARs. The specific estimators that we consider are the Bartlett estimator and the estimator proposed by Andrews and Monahan (1992). We evaluate the properties of our alternative procedures using Monte Carlo experiments in which data are generated from estimated Real Business Cycle (RBC) models. We focus on the properties of estimated impulse response functions. In our examples, the alternative procedures have better small sample properties than the standard procedure: they are associated with smaller bias, smaller mean square error and better coverage rates for confidence intervals.

The small sample properties of spectral density estimators can depend on the properties of the underlying data generating mechanism. Therefore some caution must be taken in extrapolating from our results. At a minimum our results suggest that applied econometricians should consider alternative procedures for estimating long-run identified vector autoregressions.

2 Long-Run Identifying Restrictions and VARs

There is an important literature in which researchers impose long-run identifying restrictions on VAR. For example, Gali (1999) imposes the identifying assumption that the only shock that affects the log of labor productivity, a_t, in the long run is a technology shock ε_t^z. Gali's identifying assumption corresponds to the *exclusion restriction*:

$$\lim_{j\to\infty}\left[E_t a_{t+j} - E_{t-1}a_{t+j}\right] = f\left(\varepsilon_t^z \text{ only}\right). \tag{1}$$

In addition Gali imposes the *sign* restriction that f is an increasing function of ε_t^z.

Gali imposes the exclusion and sign restrictions on a VAR to compute ε_t^z and identify its dynamic effects on macroeconomic variables. Denote the N variables in a VAR by Y_t :

$$\begin{aligned} Y_{t+1} &= B\left(L\right)Y_t + u_{t+1},\ Eu_{t+1}u_{t+1}' = V, \\ B(L) &\equiv B_1 + B_2 L + ... + B_q L^{q-1}, \\ Y_t &= \left(\begin{array}{c} \Delta \log a_t \\ \log l_t \\ x_t \end{array}\right). \end{aligned} \tag{2}$$

Here l_t denotes per capita hours worked and x_t is an additional vector of variables that may be included in the VAR. In all of our applications we assume that $q = 4$. Suppose that the fundamental economic shocks are related to u_t via the relationship:

$$u_t = C\varepsilon_t,\ E\varepsilon_t\varepsilon_t' = I,\ CC' = V, \tag{3}$$

where the first element in ε_t is ε_t^z. It is easy to verify that:

$$\lim_{j\rightarrow\infty} E[a_{t+j}|\Omega_t] - E_{t-1}[a_{t+j}|\Omega_{t-1}] = \tau\left[I - B(1)\right]^{-1} C\varepsilon_t. \tag{4}$$

Here τ is a row vector with all zeros, except unity in the first location and $B(1)$ is the sum, $B_1 + ... + B_q$. Also, E is the expectation operator and is $\Omega = \{Y_t, ..., Y_{t-q+1}\}$ is the information set.

To compute the dynamic effects of ε_t^z on the elements of Y_t we need to know $B_1, ..., B_q$ and C_1, the first column of C. The matrix, V, and the B_i's can be estimated by an ordinary least squares regression. The requirement that $CC' = V$ is insufficient to determine a unique value of C_1. Adding the exclusion and sign restrictions does uniquely determines C_1. The exclusion restriction implies that:

$$\left[I - B(1)\right]^{-1} C = \left[\begin{array}{cc} \text{number} & \underset{1\times(N-1)}{0} \\ \text{numbers} & \text{numbers} \end{array}\right].$$

The 0 matrix reflects the assumption that only a technology shock can have a long run effect on a_t. The sign restriction implies that the $(1,1)$ element of $\left[I - B(1)\right]^{-1} C$ is positive.

The standard algorithm for computing C_1 requires the matrix

$$D \equiv \left[I - B(1)\right]^{-1} C.$$

Although we cannot directly estimate D, we can estimate DD' :

$$DD' = \left[I - B(1)\right]^{-1} V \left[I - B(1)'\right]^{-1}. \tag{5}$$

The right hand side of (5) can be computed directly from the VAR coefficients and the variance-covariance matrix of the u_t. There are many D matrices consistent with (5) as well as the sign and exclusion restrictions. However, these matrices all have the same first column, D_1 (see Christiano, Eichenbaum, and Vigfusson CEV (2005)). This unique D_1 implies a unique value for C_1:

$$C_1 = [I - B(1)]\, D_1. \tag{6}$$

The right hand side of (5) is the estimate of the zero-frequency spectral density matrix of Y_t, $S_Y(0)$, that is implicit to the VAR of order q. When the true value of q is larger than the one used by the econometrician, there is reason to be concerned about the statistical properties of this estimator. A mismatch between the true value of q and the value that the econometrician assumes can arise when the data generating process is a dynamic stochastic general equilibrium model. For these types of models, Y_t may have an infinite-ordered VAR, i.e. $q = \infty$.

To understand the nature of the problem when q is misspecified, it is useful to consider a simple analytic expression closely related to results in Sims (1972). Equation (7) approximates what an econometrician who fits a misspecified VAR will find. The expression is an approximation because it assumes a large sample of data. Let $\hat{B}_1, ..., \hat{B}_q$ and $\hat{V}$ denote the estimated parameters of a VAR with q lags. Then,

$$\hat{V} = V + \min_{\hat{B}_1,...,\hat{B}_q} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[B\left(e^{-i\omega}\right) - \hat{B}\left(e^{-i\omega}\right)\right] S_Y(\omega) \left[B\left(e^{i\omega}\right) - \hat{B}\left(e^{i\omega}\right)\right]' d\omega, \tag{7}$$

where $B\left(e^{-i\omega}\right)$ is $B(L)$ with L replaced by $e^{-i\omega}$.[1] Here, B and V are the parameters of the actual VAR representation of the data, and $S_Y(\omega)$ is the associated spectral density at frequency ω.

According to (7), estimating an unconstrained VAR approximately involves choosing parameters to minimize a quadratic form in the difference between the estimated and true lag matrices. The quadratic form assigns greatest weight to the frequencies where the spectral density is the greatest. If the econometrician's VAR is correctly specified, then $\hat{B} = B$ and $\hat{V} = V$. If the econometrician estimates a VAR with too few lags relative to the true VAR, then this specification error implies that $\hat{B} \neq B$ and $\hat{V} > V$.[2] Equation (7) indicates that $\hat{B}(1)$ will be a good approximation for $B(1)$ only if $S_Y(\omega)$ happens to be relatively large in a neighborhood of $\omega = 0$. This is not something we can rely on.

Our alternative procedure for estimating long-run identified VARs involves

[1] The minimization is over the trace of the indicated integral.

[2] By $\hat{V} > V$ we mean that $\hat{V} - V$ is a positive definite matrix.

replacing the zero-frequency spectral density estimator in (5) by other estimators specifically designed for the task. There is a very large literature that analyzes the properties of alternative estimators of zero-frequency spectral density matrices. One alternative is to use the Bartlett estimator:

$$S_Y(0) = \sum_{k=-(T-1)}^{T-1} g(k)\hat{C}(k), \; g(k) = \begin{cases} 1 - \frac{|k|}{r} & |k| \leq r \\ 0 & |k| > r \end{cases}, \tag{8}$$

where, after removing the sample mean from Y_t,

$$\hat{C}(k) = \frac{1}{T} \sum_{t=k+1}^{T} Y_t Y_{t-k}'.$$

Another alternative is to use the Andrews-Monahan (1992) estimator:

$$\begin{aligned} S_Y(0) &= [I - (B_1 + ... + B_p)]^{-1} F_u(0) \left[I - (B_1 + ... + B_p)'\right]^{-1} \quad (9) \\ F_u(0) &= \sum_{k=-r}^{r} |1 - \frac{k}{r}| G(k), \quad G(k) = \frac{1}{T} \sum_{t=k}^{T} u_t u_{t-k}' \end{aligned}$$

In our alternative procedure for estimating the dynamic effect of a technology shock, we replace the estimate of $S_Y(0)$, used to construct DD' and D_1 in the standard procedure, with the estimator of $S_Y(0)$ given by either (8) or (9). We refer to these procedures as the Bartlett and Andrews-Monahan procedures, respectively. We then construct C_1 using (6).[3] In the alternative procedures C_1 still depends on the sum of the VAR coefficients even though D_1 does not. To implement (8) or (9) we use essentially all possible covariances in the data by choosing a large value of r, $r = 150$.[4] In addition, we examine the properties of our alternative procedures when we use smaller values of r.

3 The Data Generating Process

CEV (2005) show that a standard RBC model with a permanent technology shock and a labor supply shock implies the following reduced form model for

[3] The C matrix constructed with this C_1 will not have the property that $CC' = V$.

[4] The rule of always setting the bandwidth, r, equal to sample size does not yield a consistent estimator of the spectral density at frequency zero. We assume that as the sample size increases the bandwidth is increased sufficiently slowly to obtain a consistent estimator.

$Y_t = [\Delta \log a_t, \log l_t]'$:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \tag{10}$$

where ϕ_1 and ϕ_2 are scalars, θ_i is two by two matrix and ε_t is a normally distributed vector with variance-covariance matrix Ψ. The sign and exclusion restrictions are satisfied by the RBC model that CEV(2005) consider. Equation (10) implies that Y_t has an infinite-order VAR. Therefore, any finite-order VAR is misspecified.

In the *benchmark version* of the model, we use values for the exogenous shock processes obtained by CEV (2005) who maximize the Gaussian log likelihood function for Y_t using postwar U.S. data over the sample 1959QIII to 2001QIV. We summarize their results as follows:

$$\begin{aligned}\phi_1 &= 1.9439,\ \phi_2 = -0.9445,\ \Psi = \frac{1}{1000}\begin{pmatrix} 0.039137 & -0.0095548 \\ -0.0095548 & 0.073729 \end{pmatrix}\\ \theta_1 &= \begin{pmatrix} -1.926 & 0.027834 \\ -0.0066342 & -0.97012 \end{pmatrix},\ \theta_2 = \begin{pmatrix} 0.92682 & -0.027548 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

In the CKM version of the model, we assume values for the parameters equal to those used by Chari, Kehoe, and McGrattan's (CKM) (2005) in their benchmark model. CKM obtain these values using a Kalman filter framework to estimate the model. CKM impose the highly questionable identifying assumption that, up to a small measurement error, the change in government spending and net exports is a good measure of the change in technology. CEV (2005) show that there is overwhelming empirical evidence against the restrictions imposed by CKM. Despite the empirical implausibility of the CKM model, we use it as a data generating process because it provides us with an additional example to assess our alternative procedures. CKM's estimated parameter values are given by:

$$\begin{aligned}\phi_1 &= 1.9093,\ \phi_2 = -0.9114, \Psi = \frac{1}{1000}\begin{pmatrix} 0.11604 & -0.14605 \\ -0.14605 & 0.52697 \end{pmatrix}\\ \theta_1 &= \begin{pmatrix} -1.8967 & 0.041277 \\ 0.0064704 & -0.96777 \end{pmatrix},\ \theta_2 = \begin{pmatrix} 0.89971 & -0.040268 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

4 Results

To analyze the properties of our different procedures we generate multiple data sets, each of length 180 observations. We generate 1000 data sets using the benchmark model and 1000 data sets using the CKM model as the data generating mechanism. For each data set, we estimate a four-lag VAR and we estimate $S_Y(0)$ using the standard VAR-based estimator, the Bartlett estimator, and the Andrews-Monahan estimator. We then calculate three different estimates of the dynamic response of hours worked to a technology shock, as described in Section 2.

Figure 1 reports the average estimated response across all simulations (the solid line) and the true model response (the starred line). In addition, Figure 1 reports the true degree of sampling uncertainty, measured by the standard deviation of the estimated impulse response functions across the 1000 synthetic data sets (the dashed interval). To assess whether an econometrician would correctly estimate the true degree of sampling uncertainty, we proceed as follows. For each synthetic data set, and corresponding estimated impulse response function, we calculate the bootstrap standard deviation of each point in the impulse response function. Specifically, we estimate a VAR on each of the 1000 data sets that we simulate from the economic model. We use the VAR coefficients and fitted disturbances in a bootstrap procedure to generate 200 synthetic data sets, each of length 180 observations. For each of these 200 synthetic data sets, we estimate a new VAR and impulse response function. We then calculate the standard deviation of the coefficients in the impulse response functions across the 200 data sets. We use this standard deviation to construct a two standard deviation confidence interval around the estimated responses. We also calculate coverage rates for the first coefficient in the impulse response function. Specifically, we report how often, across the 1000 data sets simulated from the economic model, the econometrician's confidence intervals contain the first coefficient of the true impulse response function. In addition, Figure 1 reports the average confidence interval (the circles) that an econometrician would construct.

To help quantify the statistical properties of the different procedures Table 1 reports summary statistics for X, the estimated contemporaneous response of l_t to a technology shock.: median, mean and standard deviation of X across the 1000 synthetic data sets for each model. A different metric for assessing the different procedures is to focus on the mean square error (MSE) of X. The MSE combines information about bias and sampling uncertainty. We calculate the square root of the MSE, $\left[(1000)^{-1}\sum_{i=1}^{1000}(X_i-\xi)^2\right]^{1/2}$, for our different procedures. Here X_i denotes the value of X obtained in the i^{th} synthetic data set and ξ is the true response. Table 1 reports the value of this statistic for the different procedures relative to the MSE associated with the standard procedure. Finally we report coverage rates for X.

Consider first the results obtained using the CKM model as the data generating process. Figure 1 indicates that there is substantial upward bias associated with the standard procedure, although the bias is small relative to sampling uncertainty. Figure 1 indicates that the size of the bias associated with the Andrews - Monahan procedure is substantially smaller than the bias associated with the standard procedure. The bias associated with the Bartlett procedure is smaller still. Table 1 indicates that the bias for X falls from 0.63 using the standard procedure to 0.31 for the Andrews - Monahan procedure and to 0.08 for the Bartlett procedure. The square root of the MSE for X also falls with the modified procedures: the ratio of the MSE for the Andrews - Monahan and Bartlett procedures relative to the standard procedure is 0.91 and 0.77, respectively.

Table 1 indicates that the coverage rate for X of our alternative procedures improves relative to the standard procedure. For the standard procedure, the coverage rate for X is only 74 percent which is substantially below the nominal size of 95 percent. For the Andrews - Monahan procedure, the coverage rate is 91 percent and the coverage rate is 94 percent for the Bartlett procedure. Finally, Table 1 indicates that for both alternative procedures, setting r to 150 leads to better properties than setting r to 25 or 50.

Taken together, our results indicate that for data generated by the CKM model, our alternative procedures perform better than the standard procedure with smaller bias, smaller mean square error, and coverage rates that are closer to the nominal size. In this particular application, the Bartlett procedure outperforms the Andrews-Monahan procedure.

We now briefly discuss results when the data generating mechanism is our benchmark model. Figure 1 shows that the bias associated with the standard procedure is much smaller than when the data are generated using the CKM model. Even so, both modified procedures outperform the standard procedure. Again the bias is smaller, the MSE is reduced and coverage rates are closer to nominal sizes. Judging between the Andrews-Monahan and Bartlett procedures is more difficult. Bias is larger for the Bartlett procedure but the MSE is smaller and the coverage rate is closer to the nominal size.

5 Conclusion

We develop alternative procedures for estimating long-run identified vector autoregressions. We assess their properties relative to the standard procedure used in the literature. We evaluate the properties of the different procedures using Monte Carlo experiments where we generate data from estimated RBC models. Focusing on estimated response functions, we find that our alternative procedures have better small sample properties than the standard procedure: they are associated with smaller bias, smaller mean square error and better coverage rates for confidence intervals.

References

[1] Andrews, D.W.K., and J.C. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* 60, pp.953-966.

[2] Chari, V.V., Patrick Kehoe, and Ellen McGrattan, 2005, 'A Critique of Structural VARs Using Real Business Cycle Theory,' Federal Reserve Bank of Minneapolis Staff Report 364 July.

[3] Christiano, Lawerence J., Martin Eichenbaum and Robert Vigfusson, 2005, 'Assessing Structural VARs', manuscript, Northwestern University.

[4] Gali, Jordi, 1999, 'Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?' *American Economic Review*, 89(1), 249-271.

[5] Sims, Christopher, 1972, 'The Role of Approximate Prior Restrictions in Distributed Lag Estimation,' *Journal of the American Statistical Association*, 67(337), 169-175.

Table 1: The Contemporaneous Response of Hours Worked to a Positive Technology Shock

Procedure	r	Median	Mean	Standard Deviation	MSE*	Coverage
		CKM Model (Hours Response 0.32)				
Standard		1.03	0.95	0.68		0.74
Andrews-Monahan	25	1.16	1.04	0.81	1.17	0.76
	50	0.98	0.85	0.92	1.14	0.87
	150	0.74	0.63	0.79	0.91	0.91
Bartlett	25	0.53	0.48	0.52	0.58	0.90
	50	0.60	0.51	0.78	0.86	0.92
	150	0.44	0.40	0.71	0.77	0.94
		Our Benchmark Model (Hours Response 0.28)				
Standard		0.43	0.33	0.42		0.84
Andrews-Monahan	25	0.44	0.34	0.48	1.13	0.86
	50	0.39	0.29	0.53	1.25	0.88
	150	0.28	0.21	0.44	1.06	0.87
Bartlett	25	0.16	0.14	0.22	0.63	0.89
	50	0.19	0.15	0.35	0.89	0.92
	150	0.15	0.12	0.33	0.87	0.91

Notes: Results from Monte Carlo experiments with data simulated from two parameterizations of a RBC model.
*MSE represents the ratio of the square root of the mean squared error relative to the standard procedure

Figure 1: The Response of Hours Worked to a Positive Technology Shock