

Block Recursion and Structural Vector Autoregressions

Tao Zha*

Federal Reserve Bank of Atlanta, Atlanta, GA 30303

June 1996

Revised, October 1997

Final version, May 1998

Abstract

In applications of structural VAR modeling, finite-sample properties may be difficult to obtain when certain identifying restrictions are imposed on lagged relationships. As a result, even though imposing some lagged restrictions makes economic sense, lagged relationships are often left unrestricted to make statistical inference more convenient. This paper develops block Monte Carlo methods to obtain both maximum likelihood estimates and exact Bayesian inference when certain types of restrictions are imposed on the lag structure. These methods are applied to two examples to illustrate the importance of imposing restrictions on lagged relationships.

Key Words: Structural VAR; contemporaneously recursive blocks; identifying restrictions; likelihood; finite samples; posterior; block Monte Carlo methods.

JEL classification: C11; C15; C32; E52

* Detailed comments from four referees, the associate editor, and Arnold Zellner have led to significant improvement over earlier drafts (formerly entitled “Identification, Vector Autoregression, and Block Recursion”). The author has also benefited from discussions with and suggestions by numerous people, especially Frank Diebold, Jon Faust, Lutz Kilian, Eric Leeper, Adrian Pagan, Chris Sims, Ellis Tallman, Dan Waggoner, and Chuck Whiteman. Bryan Acree and David Petersen provided research assistance. Computation is executed in Matlab. The code is available upon request. The views expressed here are not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1. Introduction

When Sims (1980) introduced vector autoregressions (VAR) into economics, the main thrust was that VAR modeling avoids “incredible” identifying assumptions made by traditional large-scale macroeconometric models. Subsequently, the great bulk of structural VAR work has focused on contemporaneous relationships between variables or between residuals in a system of equations.¹ In a recent paper by Sims and Zha (1997), they show how to make Bayesian inference under a flat prior in both reduced-form VARs and identified VARs. In that paper, they consider various types of identifying restrictions only on contemporaneous coefficients.

There are instances, however, in which overidentification in VAR relates to lag structure as certain lags do not enter certain equations. In many empirical applications, such restrictions are not unreasonable; on the contrary, restrictions on the lag structure are necessary precisely on the ground of economic reasoning. These situations frequently stem from some block exogeneity restrictions such as the crucial small-open-economy feature in international economics or from some beliefs that certain lags do not appear in certain equations (e.g., Zellner and Palm (1974), Zellner (1985), Leeper and Gordon (1992), Sims and Zha (1995), and Bernanke, Gertler, and Watson (1997)). Failing to impose these restrictions because they may complicate statistical inference not only is economically unappealing but also may result in misleading conclusions.

This paper develops Bayesian methods that can be readily applied to economic problems that surface when overidentification in VAR relates to lag structure. The methods deal with

¹ A sample of such work includes Bernanke (1986), Blanchard and Watson (1986), Sims (1986), Blanchard (1989), Gali (1992), Gordon and Leeper (1994), Pagan and Robertson (1995), Bernanke and Mihov (1996), Eichenbaum and Evans (1995), Sims and Zha (1995), Christiano, Eichenbaum, and Evans (1996,1997), Strongin (1995), Leeper, Sims, and Zha (1996), Cochrane

situations wherein a structural model is composed of blocks that are recursive in the coefficients of *contemporaneous* variables and wherein lag structure, due to possible restrictions on lagged behaviors, may change from block to block.² The paper distinguishes between strong and weak recursive blocks in the contemporaneous coefficient matrix. In the first situation, there are no excess restrictions on contemporaneous coefficients. Thus, even if the model is overidentified by some restrictions on lagged variables, one can easily define a reduced form in which there are no lag restrictions simply by exploring the recursive block structure. The solution resembles the traditional one for fully recursive models (e.g., Zellner (1971), pp. 250-252). What is new is the development of a block Monte Carlo (MC) method.

The situation of weak recursive blocks relates to a combination of some form of block recursion and excess restrictions on contemporaneous coefficients. Because of certain restrictions on lagged coefficients, conceptual and numerical difficulties arise when an exact solution needs to be obtained. Consequently, the convenient Bayesian procedures proposed by Sims and Zha (1997) fail to apply here. Indeed, such difficulties have caused earlier researchers either to provide no error bands for impulse responses (e.g., Racette and Raynauld (1992) and Sims and Zha (1995)), or to suggest some intuitive but nonetheless infeasible iterative procedures (Dias, Machado, and Pinheiro (1996)). This paper sets out a generalized block MC method to show that it is possible to obtain maximum likelihood (ML) estimation and exact Bayesian inference with substantial computational gains. Although the framework is developed especially for situations wherein certain restrictions arise from lag structures, it includes as a

(1996), Uhlig (1997), and Bernanke, Gertler, and Watson (1997).

² The emphasis on contemporaneous block recursivity goes back to Fisher (1966, Chapter 4).

special case structural VARs with only contemporaneous restrictions discussed in Sims and Zha (1997).

To focus on deriving finite-sample properties for the class of aforementioned structural VAR models, this paper refrains from comparing the Bayesian method with other methods such as some classical ones or incorrect ones. Such a comparison has been thoroughly made by Kilian (1997) and by Sims and Zha (1997) in VARs with contemporaneous restrictions. There is little to be gained by repeating the comparison. Rather, this paper points out cases of potential misuse of Bayesian methods.

The main objective of this paper is to provide ways of exploring implications of the data that are contained in the likelihood function. Many researchers, Bayesians or non-Bayesians, would be interested in the shape of the likelihood. Since the likelihood function is a posterior probability density under a flat prior, this paper also refrains from considering non-reference (informative) priors so that the method as a scientific-reporting device appeals to as wide an audience as possible. With the shape of the likelihood, one can always quantify one's own prior knowledge in particular problems (Ingram and Whiteman (1994) and Leeper, Sims, and Zha (1996)).³

Section 2 of this paper lays out a general framework of structural VAR models. Section 3 analyzes models with strong recursive blocks, provides a block MC method that lays the foundation for the rest of the paper, discusses practical implications of these models with a brief list of applications, and points out the source of errors that occurred in previous works. Section 4

³ A widely used informative prior in VAR models is known as “the Minnesota prior” (Litterman (1986)). Sims and Zha (1996) extend such a prior to the structural VAR framework. The methods developed in this paper are also valid under the Sims and Zha prior.

studies the model by Bernanke, Gertler, and Watson (1997) to show how the block MC method can be used to eliminate certain anomalous results discussed by them. Section 5 explores models with weak recursive blocks and establishes general results that include the method of Sims and Zha (1997) and the method of Section 3 as special cases. In Section 6, the small-open-economy example of Cushman and Zha (1997) is used to show the advantage of using the model with weak recursive blocks.

2. General Setup

Consider structural VAR models of the general form:⁴

$$A(L)y(t) = \varepsilon(t), \quad t = 1, \dots, T, \quad (1)$$

where $A(L)$ is an $M \times M$ matrix of non-negative-power polynomials in lag operators with lag length p , $A(0)$ (the coefficient matrix of L^0 in $A(L)$) is nonsingular, $y(t)$ is an $M \times 1$ vector of observed variables, and $\varepsilon(t)$ is an $M \times 1$ vector of structural disturbances. Disturbance $\varepsilon(t)$ is assumed to be Gaussian with

$$E[\varepsilon(t)\varepsilon(t)'|y(t-s), s > 0] = I, \quad E[\varepsilon(t)|y(t-s), s > 0] = 0, \quad \text{all } t, \quad (2)$$

where I is the identity matrix with dimension M . Partition $A(L)$ into

$$(A_{ij}(L))(i = 1, \dots, n, j = 1, \dots, n),$$

where each element $A_{ij}(L)$ is an $m_i \times m_j$ matrix of polynomials in lag operators and

$$m_1 + \dots + m_n = M.$$

Accordingly, system (1) can be divided into a set of blocks:

⁴ For expository clarity, deterministic variables are ignored.

$$A_i(L)y(t) = \varepsilon_i(t), \quad i = 1, \dots, n, \quad \text{all } t, \quad (3)$$

where $A_i(L)$ is the matrix $(A_{i1}(L), \dots, A_{im}(L))$ and $\varepsilon_i(t)$ is an $m_i \times 1$ vector of corresponding disturbances in block i . Throughout the paper, the assumption that the lag structure is the same within a block is maintained, but identifying restrictions on the lagged coefficients in $A(L)$ are allowed so that lag structures can differ across blocks.

3. Strong Recursive Blocks

The analysis begins with models that have strong recursive blocks in the contemporaneous coefficient matrix.

Definition 1. Model (1) has strong recursive blocks in the contemporaneous coefficient matrix $A(0)$ if $A_{ij}(0) = 0$ for $i > j$ and $A_{ij}(0)$ is unrestricted for $j \geq i$.

Before proceeding, several notations are in order. Denote the block diagonal coefficient matrix of L^0 in $A(L)$ by

$$A_d(0) = \text{diag}(A_{11}(0), \dots, A_{ii}(0), \dots, A_{nn}(0)),$$

and rewrite model (1) as

$$A_d^{-1}(0)A(L)y(t) = v(t), \quad \text{all } t, \quad \text{where } v(t) = A_d^{-1}(0)\varepsilon(t). \quad (4)$$

Define

$$m_{i-} = \begin{cases} 0 & i = 1 \\ m_1 + \dots + m_{i-1} & i = 2, \dots, n \end{cases}; \quad m_{i+} = \begin{cases} m_{i+1} + \dots + m_n & i = 1, \dots, n-1 \\ 0 & i = n \end{cases}.$$

System (4) can then be divided into normalized blocks, each of which is simply the rearranged form of the original block in (3):

$$y_i(t) = C_i(L)y(t) + v_i(t), \quad i = 1, \dots, n, \quad \text{all } t, \quad (5)$$

$$C_i(L) = (\mathbf{0}_{i-}, I_i, \mathbf{0}_{i+}) - A_{ii}^{-1}(0)A_i(L), \quad (6)$$

where $\mathbf{0}_{i-}$ is the matrix of zeros with dimension $m_i \times m_{i-}$, I_i is the identity matrix with dimension m_i , $\mathbf{0}_{i+}$ is the matrix of zeros with dimension $m_i \times m_{i+}$, and $y_i(t)$ is an $m_i \times 1$ vector of observed contemporaneous variables in block i .

System (5) consists of a set of n blocks of equations. In each block i , the right-hand side of the equations contains contemporaneous variables $y_j(t)$ only for $j > i$. Normalized disturbances $v(t)$'s have the characteristic of block orthogonality:

$$E[v(t)v(t)'|y(t-s), s > 0] = \text{diag}(\Sigma_{11}, \dots, \Sigma_{ii}, \dots, \Sigma_{nn}) \quad (8)$$

$$\Sigma_{ii} = A_{ii}^{-1}(0)A_{ii}^{-1}(0)' \quad (9)$$

Obviously, framework (5) is widely used in the literature. The leading case is a class of VAR models with Choleski decompositions of the estimated covariance matrix of residuals. In these cases, each equation simply forms a block. A more sophisticated use of (5), however, concerns cases in which the contemporaneous coefficient matrix $A_{ii}(0)$ is non-recursive or lag structures differ across blocks or both.

Let k_i be the total number of right-hand-side variables per equation in i th block of (5) and rewrite (5) in matrix form:

$$\mathbf{Y}_i = \mathbf{X}_i \mathbf{C}_i + \mathbf{v}_i \quad , \quad i = 1, \dots, n \quad (10)$$

$T \times m_i$ $T \times k_i$ $k_i \times m_i$ $T \times m_i$

where \mathbf{Y}_i is a matrix of observations of contemporaneous variables, \mathbf{X}_i is a matrix of observations of lagged variables as well as contemporaneous variables (\mathbf{Y}_j 's ($j > i$)) from other blocks, \mathbf{C}_i is the matrix form of $C_i(L)$, and \mathbf{v}_i is the matrix form of $v_i(t)$ ($t = 1, \dots, T$). It can be seen from (1) that the conditional p.d.f. of $y(t)$ is

$$p(y(t)|y(t-s), s > 0) \propto |A(0)| \exp \left[-\frac{1}{2} (A(L)y(t))' (A(L)y(t)) \right].$$

Thus, the joint p.d.f. of the data $y(1), \dots, y(T)$, conditional on the initial observations of y , is proportional to

$$\begin{aligned}
& |A_d(0)|^T \exp \left[-\frac{1}{2} \sum_t (A(L)y(t))' (A(L)y(t)) \right] \\
& \propto \prod_i |A_{ii}(0)|^T \exp \left[-\frac{1}{2} \text{trace}(S_i(\mathbf{C}_i) A_{ii}(0)' A_{ii}(0)) \right] \\
& \propto \prod_i |A_{ii}(0)|^T \exp \left[-\frac{1}{2} \text{trace} \left(S_i(\hat{\mathbf{C}}_i) A_{ii}(0)' A_{ii}(0) + (\mathbf{C}_i - \hat{\mathbf{C}}_i)' \mathbf{X}_i' \mathbf{X}_i (\mathbf{C}_i - \hat{\mathbf{C}}_i) A_{ii}(0)' A_{ii}(0) \right) \right], \quad (11)
\end{aligned}$$

where

$$\hat{\mathbf{C}}_i = \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{Y}_i, \quad S_i(\mathbf{C}_i) = (\mathbf{Y}_i - \mathbf{X}_i \mathbf{C}_i)' (\mathbf{Y}_i - \mathbf{X}_i \mathbf{C}_i). \quad (12)$$

In the first line of (11), the equality $|A_d(0)| = |A(0)|$ is used because $A(0)$ is block recursive. The second line of (11) indicates that this equality is crucial for dividing the likelihood function for the whole system into separate likelihood functions for individual blocks. The property of breaking the likelihood into m blocks is related to the SUR idea introduced by Zellner (1962), when contemporaneous correlations of residuals are zero (see also Zellner (1971), pp. 250-252). Here, the analysis takes up the structural form directly. The advantage, as shown below, is that Bayesian inference can be derived conveniently.

From the third line of (11), it is clear that the concentrated likelihood for $A_{ii}(0)$ is

$$|A_{ii}(0)|^T \exp \left[-\frac{1}{2} \text{trace} \left(S_i(\hat{\mathbf{C}}_i) A_{ii}(0)' A_{ii}(0) \right) \right]. \quad (13)$$

To inform readers of the overall likelihood shape, the analysis here is restricted to only a diffuse prior on $(A_{ii}(0), \mathbf{C}_i)$. In particular, this paper uses the diffuse prior $|A_{ii}(0)|^{k_i}$ suggested by Sims and Zha (1997) for elements of $A_{ii}(0)$. As shown in the following theorem, this prior eliminates

possible discrepancies between the posterior mode and the ML estimate.⁵ The prior on all other parameters is flat. Let $\varphi(\mu; \Omega)$ denote a normal p.d.f. with mean μ and covariance matrix Ω and $\text{vec}(A)$ the vectorized column of matrix A . The theorem and algorithm, following, establish a *block* MC method.

Theorem 1. Under the diffuse prior $|A_{ii}(0)|^{k_i}$ on $(A_{ii}(0), \mathbf{C}_i)$, the joint posterior p.d.f. of $(A_{ii}(0), \mathbf{C}_i)$ is

$$p(A_{ii}(0))p(\mathbf{C}_i|A_{ii}(0)),$$

where

$$p(A_{ii}(0)) \propto (13),$$

$$p(\text{vec}(\mathbf{C}_i)|A_{ii}(0)) = \varphi\left(\text{vec}(\hat{\mathbf{C}}_i); (A_{ii}(0)'A_{ii}(0))^{-1} \otimes (\mathbf{X}_i' \mathbf{X}_i)^{-1}\right). \quad (14)$$

Furthermore, the value of $(A_{ii}(0), \mathbf{C}_i)$ at the peak of its posterior density is the ML estimate.

Proof. Under prior $|A_{ii}(0)|^{k_i}$, the likelihood (11) implies that the joint posterior p.d.f. of $(A_{ii}(0), \mathbf{C}_i)$ is

$$|A_{ii}(0)|^{T+k_i} \exp\left[-\frac{1}{2} \text{trace}\left(S_i(\hat{\mathbf{C}}_i)A_{ii}(0)'A_{ii}(0) + (\mathbf{C}_i - \hat{\mathbf{C}}_i)' \mathbf{X}_i' \mathbf{X}_i (\mathbf{C}_i - \hat{\mathbf{C}}_i)A_{ii}(0)'A_{ii}(0)\right)\right]. \quad (15)$$

Note that the second part of the exponential term in (15) is

$$\text{trace}\left((\mathbf{C}_i - \hat{\mathbf{C}}_i)' \mathbf{X}_i' \mathbf{X}_i (\mathbf{C}_i - \hat{\mathbf{C}}_i)A_{ii}(0)'A_{ii}(0)\right) =$$

⁵ Section 5 provides an interpretation of prior $|A_{ii}(0)|^{k_i}$.

$$\text{vec}(\mathbf{C}_i - \hat{\mathbf{C}}_i)' \left(A_{ii}(0)' A_{ii}(0) \otimes \mathbf{X}_i' \mathbf{X}_i \right) \text{vec}(\mathbf{C}_i - \hat{\mathbf{C}}_i). \quad (16)$$

Since (15) does not involve \mathbf{C}_i elsewhere except for term (16), the p.d.f. of \mathbf{C}_i has the exact form of the normal p.d.f. (14). Integrating out \mathbf{C}_i in (15) produces the term $|A_{ii}(0)|^{-k_i}$ so that the marginal posterior p.d.f. of $A_{ii}(0)$ is the same as the concentrated likelihood (13). Consequently, the value of $(A_{ii}(0), \mathbf{C}_i)$ at the peak of its posterior density is the same as the ML estimate.

Q.E.D.

Given $A_{ii}(0)$ and $C_i(L)$, it follows from (6) that $A_i(L)$ can be calculated as

$$A_i(L) = A_{ii}(0) \left((\mathbf{0}_{i-}, I_i, \mathbf{0}_{i+}) - C_i(L) \right), \quad i = 1, \dots, n. \quad (17)$$

Denote a function of $A(L)$ by $f(A(L))$. By Theorem 1, Bayesian inference of $f(A(L))$ can be obtained through MC samples that are generated block by block. The algorithm, following, formalizes the procedure.

Algorithm 1. The numerical procedure involves the following steps:

For $i = 1, \dots, n$,

- (a) generate samples of $A_{ii}(0)$ by drawing from marginal posterior p.d.f. (13);
- (b) conditional on drawn $A_{ii}(0)$'s, sample \mathbf{C}_i from Gaussian distribution (14);
- (c) given MC samples of $(A_{ii}(0), \mathbf{C}_i)$, calculate samples of $A_i(L)$ by (17);

End;

- (d) calculate $f(A(L))$'s from MC samples of $A(L)$;
- (e) use these samples to compute the marginal Bayesian posterior probability interval for each element of $f(A(L))$.

The essential part of Algorithm 1 involves steps (a)-(c); they can be all done block by

block. The Bayesian method of Sims and Zha (1997) applies only to situations whereby lags are unrestricted. In this case, the whole system can be regarded as one single block and thus Theorem 1 holds for this whole system. This is exactly what Sims and Zha's procedure does. But if lag structures differ across blocks, Theorem 1 is no longer valid for the whole system as one single block. This point, obvious though it might seem, has not been recognized by all researchers. Indeed, when certain restrictions are imposed on lags, the common practice is to use the standard RATS procedure of Doan (1992) or Sims and Zha's method to generate MC draws of $A(L)$ (this amounts to using Theorem 1 to the whole system). For each draw of $A(L)$, the lag restrictions are then imposed.⁶ Such a procedure is incorrect because it fails to take account of the restrictions on lagged variables when draws of $A(L)$ are made.

The development described in Theorem 1 and Algorithm 1 lays out the foundation for the rest of the paper. In many applications, there are no excess restrictions on elements in $A_{ii}(0)$ so that there is a one-one mapping between $A_{ii}(0)$ and Σ_{ii} in (9).⁷ In this case, one can transform $A_{ii}(0)$ into Σ_{ii} in likelihood (11) according to (9). Under the improper prior $|\Sigma_{ii}|^{-\frac{k_i}{2}}$ for Σ_{ii} , the marginal posterior p.d.f. of Σ_{ii} becomes⁸

⁶ See, for example, Leeper and Gordon (1992). The author is very grateful to Eric Leeper for bringing out this point.

⁷ Examples are numerous. To sample a few, Pagan (1993) argues that certain disaggregated data do not enter as explanatory variables a set of equations that describe the aggregate behavior. Levy, Dutta, and Bergen (1996) assume that the wholesale prices of fruits do not affect their spot prices both contemporaneously and through lags. Leeper and Gordon (1992) examine the liquidity effects under the assumption that money is exogenous. Genberg, Salemi, and Swoboda (1987) treat Switzerland explicitly as a small open economy that has little influence on the foreign economy.

⁸ Prior $|\Sigma_{ii}|^{-\frac{k_i}{2}}$ is a diffuse Wishart p.d.f. in the sense of Geisser (1965) and in a univariate case is

$$|\Sigma_{ii}^{-1}|^{\frac{T}{2}} \exp\left[-\frac{1}{2} \text{trace}(S_i(\hat{\mathbf{C}}_i)\Sigma_{ii}^{-1})\right]. \quad (18)$$

From (9) and (11) it can be seen that (18) is the concentrated likelihood function for Σ_{ii} .

The p.d.f. (18) is of inverted Wishart form and Σ_{ii}^{-1} has the following Wishart distribution:⁹

$$\text{Wishart}(S_i^{-1}(\hat{\mathbf{C}}_i), T - m_i - 1, m_i). \quad (19)$$

MC samples of $A_{ii}(0)$ can be generated by first drawing Σ_{ii}^{-1} from (19) and then transforming drawn Σ_{ii}^{-1} back to $A_{ii}(0)$ according to (9). Bayesian inference of $f(A(L))$ can then be computed by following steps (b)-(e) in Algorithm 1.

Although Theorem 1 lays out the straightforward method, the impulse responses in previous works were provided either without error bands (Genberg, Salemi, and Swoboda (1987)), or with asymptotic confidence bands (Levy, Dutta, and Bergen (1996)), or with bands incorrectly computed.¹⁰ Not until recently has Algorithm 1 been used to generate correct error bands for impulse responses.¹¹

The strong recursive structure defined in Definition 1 does not exclude situations in which

simply an ignorance prior discussed in Leamer (1978, pp. 78-84). It is implied by prior $|A_{ii}(0)|^{k_i}$ for $A_{ii}(0)$ except for a Jacobian term. When $A_{ii}(0)$ is upper triangular, for example, the Jacobian is $|\partial\Sigma_{ii}^{-1}/\partial A_{ii}(0)| = 2^{m_i} \prod_{j=1}^{m_i} a_{jj}^j$, where a_{jj} 's are the diagonal elements of $A_{ii}(0)$. In practice, ignoring this Jacobian would not make much difference on the results as sample size T is in general considerably larger than m_i . For detailed discussion of these priors, see Sims and Zha (1996).

⁹ See pp. 389-396 in Zellner (1971).

¹⁰ It is well understood that asymptotic confidence bands perform very badly in small samples (see, for example, Kilian (1997)). Sims and Zha (1997) show examples where incorrect Bayesian procedures may lead to mistaken inference.

there are excess restrictions on $A_{ii}(0)$.¹² In this case, there is no one-one relationship between $A_{ii}(0)$ and Σ_{ii} . Thus, it makes no sense to draw Σ_{ii}^{-1} from (19) because $A_{ii}(0)$ can no longer be recovered uniquely from Σ_{ii} .¹³ It is probably due to this difficulty that error bands for the impulse responses were not provided in some previous works (Racette and Raynauld (1992)). One can, however, use the idea of the weighted MC method of Sims and Zha (1997) and apply it to each of the blocks that have no one-one relation between $A_{ii}(0)$ and Σ_{ii} . Specifically, $A_{ii}(0)$ can be drawn from the asymptotic Gaussian distribution approximated by the second-order Taylor expansion of the logarithm of (13) at its peak and then repeat steps (b)-(d) in Algorithm 1. Prior to step (e) in Algorithm 1, each sample of draws of $f(A(L))$ is weighted by the ratio of the product of (13)'s for these blocks to the product of those approximate Gaussian p.d.f.'s.¹⁴

4. Effects of Oil Price Shocks

This section provides a concrete example of how the method proposed in the last section can change the results established in the existing literature. Specifically, the five-variable VAR model studied by Bernanke, Gertler, and Watson (1997) is examined. The five variables are the federal funds rate (FFR), the PPI crude petroleum price index (Poil), the spot market price index

¹¹ For applications of the block MC method developed here, see, for example, Leeper (1997).

¹² A published paper by Racette and Raynauld (1992) is an example.

¹³ Previous works nonetheless generated MC draws of Σ_{ii}^{-1} directly from a Wishart, replaced $S_i(\hat{\mathbf{C}}_i)$ by drawn Σ_{ii} , and then obtained $A_{ii}(0)$ that maximizes (13) (e.g., Canova (1991) and Gordon and Leeper (1994)). But the draws of Σ_{ii} so obtained fail to take account of the excess restrictions on $A_{ii}(0)$.

¹⁴ Of course, this weighted method is valid also for cases in which there is a one-one mapping between $A_{ii}(0)$ and Σ_{ii} . In particular, if such a nonlinear mapping is complicated and requires some non-trivial computing time to transform Σ_{ii} back to $A_{ii}(0)$, the weighted procedure may be

of all commodities from the Commodity Research Bureau (Pcm), the gross domestic product (GDP) deflator (P), and real GDP (y). All variables are in logarithm except FFR, which is expressed in percentage point. All data are monthly and for the sample period 1967:1-1997:3.¹⁵

In Bernanke, Gertler, and Watson (1997), only contemporaneous restrictions are imposed and $\{FFR, Poil, Pcm, P, y\}$ is of upper-triangular order. This implies that Poil does not affect macroeconomic variables $\{Pcm, P, y\}$ contemporaneously but is affected by these variables both contemporaneously and through lags. The first column of graphics in Figure 1 displays the three 48-month impulse responses studied by Bernanke, Gertler, and Watson: the responses of output (y), the general price level (P), and the interest rate (FFR) to a one-standard-deviation shock in oil price (Poil). In Figure 1, solid lines are the ML-estimated responses; intervals between the two dashed lines contain .68 probability as one-standard-deviation error bands.¹⁶ Clearly, a shock to Poil leads to subsequent *decline* in general price level (P) while the dynamic effects on both output (y) and the interest rate (FFR) are *very small*. Bernanke, Gertler, and Watson (1997) regard these responses as “anomalous” or “unsatisfactory” because “the conventional wisdom” is

efficient compared to the procedure with direct draws from Wishart.

¹⁵ Following Bernanke, Gertler, and Watson (1997), the monthly GDP deflator and GDP are interpolated. The VAR model is estimated with a constant and six lags.

¹⁶ All error bands throughout this paper were generated with 5000 MC draws. In dynamic multivariate models, a parameter/observation ratio is typically high. In the Bernanke, Gertler, and Watson model, the ratio is 201/363. A high parameter/observation ratio tends to make error bands rather wide (Zellner (1985)). In a recent paper, Sims and Zha (1996) formulated an informative prior for the parameters of structural VAR models. The prior effectively reduces the number of free parameters and the widths of probability bands (see Leeper, Sims, and Zha (1996) for applications). Although this paper follows the standard practice in the literature of focusing only on the likelihood with a flat prior, the prior proposed by Sims and Zha seems successful in dealing with problems associated with a high parameter/observation ratio. In addition, all algorithms set out in this paper are valid for Bayesian VAR models under the Sims and Zha prior.

that oil price shocks should lead to a rise, not a decline, in the price level and should have “a significant (not small) and a priori plausible reduced-form impact on the economy.”

If one believes in this conventional wisdom, these results are not promising. The real question then is what is the cause of these “unsatisfactory” results. The analysis here argues that they may stem from the identifying restrictions placed by Bernanke, Gertler, and Watson. Movements in oil prices tended to be unrelated to the rest of the economy and, as Bernanke, Gertler, and Watson correctly point out, “were arguably exogenous, reflecting a variety of developments both in the Middle East and in the domestic industry.” Therefore, it makes economic sense to impose the exogeneity restriction on P_{oil} which is not present in the original model of Bernanke, Gertler, and Watson.

Suppose one imposes the exogeneity restriction on P_{oil} . Let the first block of equations contain all variables $\{FFR, P_{cm}, P, y\}$ and the second block contain only P_{oil} both contemporaneously and in lags. Since dynamic responses to P_{oil} shocks are invariant to how the first contemporaneous block $A_{11}(0)$ is triangularized, $\{FFR, P_{cm}, P, y\}$ is ordered to be upper triangular (i.e., $A_{11}(0)$ is upper triangular). Algorithm 1 was used to generate the impulse responses (solid lines) and .68 probability bands (dashed lines). The results of interest are reported in the second column of Figure 1. These results are consistent with the conventional wisdom. Following a shock in oil prices, the interest rate rises, reflecting the Federal Reserve’s contractionary action against possible future inflation. Consequently, output (y) declines considerably. The general price level also rises, but not as much as it would have if the Federal Reserve had not raised the interest rate in response to P_{oil} shocks. Note that the effects on the economy when the exogeneity restriction is imposed (the second column of Figure 1) are in magnitude larger than those when the restriction is ignored (the first column of Figure 1). This

example illustrates that using the method developed here one can obtain credible results if one imposes economically sensible restrictions.

5. Weak Recursive Blocks

So far, the discussion has been concerned with cases in which all elements of submatrix $A_{ij}(0)$ ($j > i$) are unrestricted. In this section, the discussion turns to a more general situation in which a strong recursive structure is insufficient to identify certain behaviors in the actual economy. This situation arises especially in the recent literature of policy analysis in which some restrictions on $A_{ij}(0)$ ($j > i$) are required to identify monetary policy (e.g., Sims and Zha (1995), Bernanke, Gertler, and Watson (1997), and Cushman and Zha (1997)). This section develops a generalized block MC method for this situation and shows that both the Bayesian procedure proposed by Sims and Zha (1997) and the block MC method in Section 3 are special cases.

The analysis begins with the following definition:

Definition 2. Model (1) has weak recursive blocks in the contemporaneous coefficient matrix $A(0)$ if $A_{ij}(0) = 0$ for $i > j$ and if there are further linear restrictions on some elements in $A_{ij}(0)$ for $j > i$.

When some restrictions on both lags and $A_{ij}(0)$ for $j > i$ are in place, neither the Bayesian procedure of Sims and Zha (1997) nor the block MC procedure developed so far works for this situation. The procedure of Sims and Zha does not work because, as discussed in Section 3, lag structures differ across blocks. The block MC method does not work here simply because the OLS estimate $\hat{\mathbf{C}}_i$ in (12) takes no account of restrictions on elements in $A_{ij}(0)$ for $j > i$ and is thus no longer the ML estimate or Bayesian posterior mode for \mathbf{C}_i .

To derive a new method, first define

$$F(L) = A(0) - A(L).$$

In agreement with the divided blocks in (3), partition $F(L)$ into a column vector of n block sub-matrices $F_i(L)$'s ($i = 1, \dots, n$) so that

$$F_i(L) = A_i(0) - A_i(L). \quad (20)$$

Hence, (3) can be rearranged as

$$A_i(0)y(t) = F_i(L)y(t) + \varepsilon_i(t), \quad i = 1, \dots, n, \quad \text{all } t. \quad (21)$$

To write (21) in compact matrix form, let \mathbf{W}_i be a $T \times q_i$ matrix of observations on the right-hand side of (21), where

$$q_i = k_i - m_{i+1} - \dots - m_n.$$

Further, let \mathbf{F}_i be the $q_i \times m_i$ matrix form of $F_i(L)$ corresponding to \mathbf{W}_i , and \mathbf{E}_i be the $T \times m_i$ matrix form of $\varepsilon_i(t)$ ($t = 1, \dots, T$). Thus the matrix version of (21) becomes:

$$\mathbf{Y} \quad A_i(0)' = \mathbf{W}_i \quad \mathbf{F}_i + \mathbf{E}_i, \quad (22)$$

$T \times M$ $M \times m_i$ $T \times q_i$ $q_i \times m_i$ $T \times m_i$

where $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)$. Note that $A_i(0) = (A_{i1}(0), \dots, A_{in}(0))$, where $A_{ij}(0) = \mathbf{0}$ for $j < i$.

Brute-force estimation of parameters $(A_i(0), \mathbf{F}_i)$ is to use some iterative procedure, an idea that occurred independently to a number of researchers (Dias, Machado, and Pinheiro (1996)).

Specifically, such a procedure involves the following steps: initiate the values of $A_i(0)$; with

initiated value $\mathbf{Z}_i = \mathbf{Y}A_i(0)'$, compute OLS estimate $\hat{\mathbf{F}}_i = \left(\mathbf{W}_i' \mathbf{W}_i \right)^{-1} \mathbf{W}_i' \mathbf{Z}_i$; calculate the

covariance matrix of reduced-form residuals for the whole system; replace $S_i(\hat{\mathbf{C}}_i)$ in (13) by this

covariance matrix and treat the whole system as one single block to find $A(0)$ that maximizes

(13); repeat this procedure using the new value of $A(0)$ until convergence is established.

Intuitive though this iterative procedure might be, it has little value in practice because, for each iteration, one has to solve a maximization problem for $A(0)$ — a very expensive task computationally. To do inference, this procedure is almost infeasible for the typical VAR systems used in the literature.

The following theorem and algorithm establish the generalized block Monte Carlo method that proves easy and inexpensive to implement. For $i = 1, \dots, n$, define

$$\mathbf{N}_i = \left(\mathbf{W}_i' \mathbf{W}_i \right)^{-1} \mathbf{W}_i' \mathbf{Y} , \quad (23)$$

$$V_i(\mathbf{N}_i) = (\mathbf{Y} - \mathbf{W}_i \mathbf{N}_i)' (\mathbf{Y} - \mathbf{W}_i \mathbf{N}_i) . \quad (24)$$

Theorem 2. Under the flat prior on $(A_i(0), \mathbf{F}_i)$, the joint posterior p.d.f. of $(A_i(0), \mathbf{F}_i)$ is

$$p(A_i(0))p(\mathbf{F}_i|A_i(0)),$$

where

$$p(A_i(0)) \propto |A_i(0)|^T \exp \left[-\frac{1}{2} \text{trace}(V_i(\mathbf{N}_i) A_i(0)' A_i(0)) \right], \quad (25)$$

$$p(\text{vec}(\mathbf{F}_i)|A_i(0)) = \varphi \left(\text{vec}(\mathbf{N}_i A_i(0)'); I_i \otimes \left(\mathbf{W}_i' \mathbf{W}_i \right)^{-1} \right) . \quad (26)$$

Furthermore, the value of $(A_i(0), \mathbf{F}_i)$ at the peak of its posterior density is the ML estimate.

Proof. By (22), the expression of likelihood function (11) can be rearranged in a different form as follows:

$$\begin{aligned} & |A_d(0)|^T \exp \left[-\frac{1}{2} \sum_t (A(L)y(t))' (A(L)y(t)) \right] \\ & \propto \prod_i |A_{ii}(0)|^T \exp \left[-\frac{1}{2} \text{trace} \left((\mathbf{Y}_i A_i(0)' - \mathbf{W}_i \mathbf{F}_i)' (\mathbf{Y}_i A_i(0)' - \mathbf{W}_i \mathbf{F}_i) \right) \right] \end{aligned}$$

$$\propto \prod_i |A_{ii}(0)|^T \exp \left[-\frac{1}{2} \text{trace} \left(V_i(\mathbf{N}_i) A_i(0)' A_i(0) + (\mathbf{F}_i - \mathbf{N}_i A_i(0))' \mathbf{W}_i' \mathbf{W}_i (\mathbf{F}_i - \mathbf{N}_i A_i(0)) \right) \right]. \quad (27)$$

Under the flat prior on $(A_i(0), \mathbf{F}_i)$, (27) is also the joint posterior p.d.f. of $(A_i(0), \mathbf{F}_i)$ for all i .

Furthermore, $(A_i(0), \mathbf{F}_i)$ is independent of all other parameters $(A_j(0), \mathbf{F}_j)$ for $j \neq i$, and its posterior p.d.f. is

$$|A_{ii}(0)|^T \exp \left[-\frac{1}{2} \text{trace} \left(V_i(\mathbf{N}_i) A_i(0)' A_i(0) + (\mathbf{F}_i - \mathbf{N}_i A_i(0))' \mathbf{W}_i' \mathbf{W}_i (\mathbf{F}_i - \mathbf{N}_i A_i(0)) \right) \right]. \quad (28)$$

Note that term $V_i(\mathbf{N}_i)$ does not contain any parameter. From (28) it is clear that the marginal posterior p.d.f. for $A_i(0)$ has form (25) and the conditional posterior p.d.f. of \mathbf{F}_i is Gaussian in form as expressed in (26).

Since the likelihood and posterior of $(A_i(0), \mathbf{F}_i)$ are the same, the ML estimate is the value of $(A_i(0), \mathbf{F}_i)$ at the peak of the posterior p.d.f.

Q.E.D.

Theorem 2 lays out the theoretical ground on which the computation of exact Bayesian inference of $f(A(L))$ can be executed. The following algorithm sets out the numerical procedure.

Algorithm 2. The procedure for computing exact finite-sample inference of $f(A(L))$ is composed of the following steps:

For $i = 1, \dots, n$,

- (a) draw $A_i(0)$ from the asymptotic Gaussian approximated by the second-order Taylor expansion of the logarithm of (25) at its peak;
- (b) conditional on drawn samples of $A_i(0)$, generate samples of \mathbf{F}_i from Gaussian

distribution (26);

(c) given generated samples of $(A_i(0), \mathbf{F}_i)$, calculate samples of $A_i(L)$ by (20);

End;

(d) calculate samples of $f(A(L))$ accordingly and then weight each sample of $f(A(L))$ by the ratio of the product of (25)'s ($i = 1, \dots, n$) to the product of the approximate Gaussian

p.d.f.'s ($i = 1, \dots, n$);

(e) use these weighted samples to compute the marginal Bayesian posterior probability interval for each parameter of $f(A(L))$.

As in Algorithm 1, crucial steps (a)-(c) in Algorithm 2 require computation only block by block. Again, when identifying restrictions are imposed only on contemporaneous coefficients, one can treat the whole system as one block and thus the procedure considered by Sims and Zha (1997) is a special case of Theorem 2. Moreover, the following theorem shows that the block MC method developed in Section 3 is also a special case.

Theorem 3. If strong block i is recursive contemporaneously, the generalized block MC of Theorem 2 attains the same estimation and inference of the block parameters as does the block MC method of Theorem 1.

Proof. Denote the submatrix of the last m_{i+} columns in $A_i(0)$ by $A_{i+}(0)$ (with dimension $m_i \times m_{i+}$) so that

$$A_i(0) = (\mathbf{0}_{i-}, A_{ii}(0), A_{i+}(0)).$$

The parameter space in Theorem 2 is

$$(A_{ii}(0), A_{i+}(0), F_i(L)). \quad (29)$$

The parameter space in Theorem 1 is

$$(A_{ii}(0), C_i(L)). \quad (30)$$

Since all elements in $A_{i+}(0)$ are unrestricted by assumption, parameter matrices (29) and (30) have a one-one mapping through the relation:

$$C_i(L) = -A_{ii}^{-1}(0)A_{i+}(0) + A_{ii}^{-1}(0)F_i(L). \quad (31)$$

Relation (31) is derived from (6) and (20).

To prove this theorem, it is sufficient to show that after parameter space (29) is transformed to (30), posterior p.d.f. (28) of parameter matrix (29) is the same as posterior p.d.f. (15) of parameter matrix (30). First, note that since (5) and (21) are simply two different arrangements of the same block of equations, likelihood functions (11) and (27) are identical. Now, given $A_{ii}(0)$, (31) implies that the mapping between $C_i(L)$ and $(A_{i+}(0), F_i(L))$ is a simple linear transformation, and the Jacobian of the transformation of $(A_{i+}(0), F_i(L))$ into $C_i(L)$ is $|A_{ii}(0)|^{k_i}$. Hence, the implied posterior p.d.f. of parameter matrix (30) from the posterior distribution of (29) is simply posterior p.d.f. (28) multiplied by Jacobian term $|A_{ii}(0)|^{k_i}$. The result leads to the exact form of (15) because likelihood functions (11) and (27) are identical.

Q.E.D.

Theorem 3 offers another interpretation of using $|A_{ii}(0)|^{k_i}$ as a diffuse prior for $A_{ii}(0)$ in Section 3. Note that parameter $C_i(L)$ is of reduced form. When specifying a prior in reduced-form parameter space $(A_{ii}(0), C_i(L))$, prior $|A_{ii}(0)|^{k_i}$ is simply to take account of the Jacobian so that when $(A_{ii}(0), C_i(L))$ is transformed to structural parameter space $(A_{ii}(0), A_{i+}(0), F_i(L))$, the prior on $A_{ii}(0)$ is flat.

Although Theorem 2 offers the method that is more general than the one in Theorem 1,

there are situations in which Theorem 1 is strictly preferred. For instance, when $A(L)$ is composed of strong recursive blocks and there is no excess restriction on $A_{ii}(0)$, inference of parameters $(A_{ii}(0), \mathbf{C}_i)$ is computationally faster because it involves neither a maximization problem nor weighted sampling. Of course, given modern computational capacity, the difference in computing time is diminishing.

The bulk of identified VAR work, when restrictions are imposed, often identifies blocks of equations without distinct behavioral interpretations for each individual equation within a block.¹⁷ Within the block, $A_{ii}(0)$ can be arbitrarily ordered through orthonormal transformation. Dynamic responses of all variables in the whole system to a shock outside block i are invariant to how $A_{ii}(0)$ is orthonormally transformed in the case of a strong recursive structure with a one-one relation between $A_{ii}(0)$ and Σ_{ii} . This result holds because each block becomes a subsystem. One special case of such an orthonormal transformation is triangularization, which has been used widely in the existing literature. Nonetheless, when the one-one relation between $A_{ii}(0)$ and Σ_{ii} does not hold or the model has a weak recursive structure as defined in Definition 2, whether the invariance result holds is still unknown. The following theorem establishes the result that applies to both strong and weak recursive structures.

Theorem 4. Let the impulse response of $y_q(t+s)$ to shock $\varepsilon_r(t)$ be defined as

$$\phi_{qr}(s) = \frac{\partial y_q(t+s)}{\partial \varepsilon_r(t)}, \quad (32)$$

where $q = 1, 2, \dots, M$, $r \leq m_{i-}$ or $r \geq m_{i+}$, and $s \geq 0$. Assume that there may be linear restrictions

¹⁷ See Leeper, Sims, and Zha (1996) for a thorough review on this point. See also Sims and Zha

on some elements in $A_{kj}(0)$ for $j \geq k$ and $k \neq i$, but $A_{ij}(0)$ is unrestricted for $j > i$. Let

$\Phi_{\bullet r}(s) = (\phi_{qr}(s))$ ($q = 1, 2, \dots, M$) be an $M \times 1$ matrix. Then the values of $\Phi_{\bullet r}(s)$ are invariant to orthonormal transformation of $A_{ii}(0)$.

Proof. Denote

$$R = \begin{pmatrix} I_{i-} & 0 & 0 \\ 0 & R_{ii} & 0 \\ 0 & 0 & I_{i+} \end{pmatrix},$$

where R_{ii} is an $m_i \times m_i$ orthogonal matrix, I_{i-} is the $m_{i-} \times m_{i-}$ identity matrix, and I_{i+} is the $m_{i+} \times m_{i+}$ identity matrix. Orthonormal transformation of $A_{ii}(0)$ is equivalent to pre-multiplying $A_{ii}(0)$ by R_{ii} . Because $A_{ij}(0)$ is unrestricted for $j \geq i$ and the same lag structure holds within block i , such a transformation amounts to arriving at the transformed system that implies the same time-series properties for the data:

$$\tilde{A}(L)y(t) = \varepsilon(t), \quad (33)$$

where

$$\tilde{A}(L) = RA(L). \quad (34)$$

Define $B(L) = I - A^{-1}(0)A(L)$ and let B_j be the coefficient matrix corresponding to L^j ($j = 1, \dots, p$) where, recall, p is the lag length. Impulse response matrix $\Phi(s)$ ($M \times M$) is defined as $(\phi_{gh}(s))$ ($g, h = 1, \dots, M$). Hence $\Phi(s)$ lies in the linear space spanned by the

(1996) and Christiano, Eichenbaum, and Evans (1997).

elements in a subset of $\left\{ \left(\prod_{h=1}^j B_h^{k_h} \right) A^{-1}(0) \right\}_{j=1, \dots, p}^{k_h \geq 0}$, where $B_h^{k_h}$ is matrix B_h raised to power k_h .

Similarly, let $\tilde{B}(L) = I - \tilde{A}_0^{-1}(L)\tilde{A}(L)$ and \tilde{B}_j be the coefficient matrix corresponding to L^j . Since R is an orthogonal matrix, (34) implies that $\tilde{A}^{-1}(0) = A^{-1}(0)R$ and thus $\tilde{B}_j = B_j$ for $j = 1, \dots, p$. Hence, impulse response matrix $\tilde{\Phi}(s)$ from system (33) lies in the linear space

spanned by the elements in a subset of $\left\{ \left(\prod_{h=1}^j B_h^{k_h} \right) A^{-1}(0)R \right\}_{j=1, \dots, p}^{k_h \geq 0}$.

Note that post-multiplying $M \times M$ matrix $\left(\prod_{h=1}^j B_h^{k_h} \right) A^{-1}(0)$ by R affects only the columns of $\left(\prod_{h=1}^j B_h^{k_h} \right) A^{-1}(0)$, where the location of these columns is indexed by m_i . Since r is outside block i (i.e., $r \leq m_{i-}$ or $r \geq m_{i+}$), it follows that $\Phi_{\bullet, r}(s) = \tilde{\Phi}_{\bullet, r}(s)$.

Q.E.D.

Theorem 4 provides a formal justification for triangularizing certain blocks in both strong and weak recursive structures when distinct interpretations for individual equations within a block are not required. Much of structural VAR work focuses on identifying the effects of a monetary policy shock and thus contemporaneous coefficient matrices in other blocks of equations often take on a triangular form (Leeper, Sims, and Zha (1996)). Theorem 4 ensures that the effects of a policy shock are invariant to such a triangular form, even in the case of a weak recursive structure.

As researchers become more aware of using a priori restrictions based on economic arguments, weak recursive structures will prove increasingly useful. One recent example is Sims and Zha (1995) and Bernanke, Gertler, and Watson (1997), who examine the effects of

systematic monetary policy under different types of rules.¹⁸ Weak recursive blocks often arise from these models. Due to the lack of available tools, however, error bands were not always provided for the estimated impulse responses. Another example relates to the area of international economics, where a crucial restriction is that a small open economy has little impact on the “rest of the world.” The next section reviews such an example.

6. Small Open Economy

In this section, the generalized block MC method is applied to the small open economy example studied by Cushman and Zha (1997). The purpose is to show what kinds of anomalous results one would obtain if a small open economy were not adequately taken into account. A nine-variable VAR model is considered. Canada is the home country here and is treated as a small open economy relative to the U.S. (foreign country) economy. There are five home-country variables: M1, the 3-month Treasury bill rate (R), the exchange rate (the U.S. dollar price of the Canadian dollar, Exc), the consumer price index (P), and industrial production (y). Four foreign variables are included in the model: the IMF world commodity export price (Wxp^*), the U.S. federal funds rate (R^*), the U.S. consumer price index (P^*), and U.S. industrial production (y^*).¹⁹ All data are monthly and are the same as in Cushman and Zha, with their sample period 1974:1-1993:12. All variables are in logarithm except interest rates that are expressed in percentage point.

The identification approach used here follows exactly what Cushman and Zha assumed.

¹⁸ Some of these policy rules under study imply that the Federal Reserve engages in interest rate targeting rather than a rule of responding to the state of the economy. Bernanke, Gertler, and Watson also assume that the federal funds rate does not enter certain equations in the system.

¹⁹ Cushman and Zha (1997) also include two additional variables (exports and imports) to

Specifically, the identification treats Canada as a small open economy and assumes that the Canadian variables do not enter as explanatory variables the equations within the foreign-country block, both contemporaneously and through lags. There are no other restrictions on lags. For contemporaneous restrictions, foreign variables $\{Wxp^*, R^*, P^*, y^*\}$ simply take on an upper triangular form. As for the Canadian economy, there are four distinct behaviors that are identified. The first is the money demand equation that takes up the functional form $M1-P=y-\alpha R$, where α is a coefficient.²⁰ The second is the Canadian monetary policy reaction function (often called “the money supply equation”), which allows the Bank of Canada (the central bank in Canada) to respond to all other variables except output (y and y^*) and price (P and P^*). This exclusion restriction on prices and output is justified because the data on output and the general price level are unobservable within the month. The third concerns the equation reflecting a financial market in which the exchange rate responds to all variables. The fourth concerns a block of equations specifying a production sector composed of $\{P, y\}$. All other contemporaneous variables are excluded from this block on the basis that these variables are probably related to home production only through lags. In this block, each individual equation is not identified and, consequently, variables $\{P, y\}$ are ordered to upper triangular. By Theorem 4, this triangularization has no effect on the dynamic responses to a monetary policy shock.

Clearly, this identification leads to weak recursive blocks.

examine the effect on the bilateral trade between Canada and the U.S.

²⁰ This is not the place to debate whether or not this restriction imposed by Cushman and Zha fits to economists’ a priori beliefs. One should note, however, that this functional form is implied by many dynamic stochastic general equilibrium models, in sharp contrast to the “long-run” money demand equation which relies on reduced-form empirical evidence. See, for example, Blanchard and Fisher (1989, p. 513), McCallum (1989, pp. 35-43), Leeper and Sims (1994), and Sims and

The marginal processes for individual variables in this structural model are of finite-order ARMA form but in a restricted manner implied by the identifying restrictions (Zellner and Palm (1974)). Although these ARMA processes may be useful for predictions, structural analysis on, say, the effects of a monetary policy shock requires one to work directly on the structural model as a whole. This point is well articulated by Zellner and Palm (1974). For this nine-variable system, therefore, the generalized block MC method (Algorithm 2) is employed to compute dynamic responses to structural shocks.

Figure 2 displays several responses of interest with .68 probability bands.²¹ These are the dynamic responses to a *contractionary* monetary policy shock. Following this contractionary shock, the money stock (M1) falls, the interest rate (R) rises initially (the liquidity effect), the exchange rate responds positively for the first year (the exchange rate effect), and both the general price level (P) and output (y) fall (no price or output puzzle). These results are consistent with a priori beliefs about the effects of monetary policy shocks in a small open economy; no anomalies produced in previous empirical studies occur in the results.

Natural questions are: How important is it to impose the small-open-economy restriction? What sorts of mistakes would one make by simply following the convenient method developed

Zha (1995).

²¹ ML estimation for the entire nine-variable system took 14.10 seconds on a Pentium II/266 machine. The computation took advantage of contemporaneously recursive blocks. If one had treated the whole system as one block, computing time would have increased two-fold for this example. In general, the more blocks a system is broken into, the more gain one will obtain in computing time. As for this nine-variable system, computation took 35.45 seconds per 1000 MC draws. To take account of possible fat tails of the true distribution of $A_i(0)$, the asymptotic Gaussian in step (a) of Algorithm 2 is replaced by a t -distribution with 9 degrees of freedom (Geweke (1989)). Alternatively, when the distribution of $A_i(0)$ is very non-Gaussian, the Metropolis method proposed by Waggoner and Zha (1997) can be used.

in the previous work (e.g., Sims and Zha (1997)) without imposing such a restriction? Table 1 reports some anomalous results when the exogeneity restriction is absent from the model. It displays the variance decompositions for U.S. variables $\{R^*, P^*, y^*\}$ that are attributed to all shocks emanating from the Canadian economy at various time horizons, along with the percentages of these decompositions due to Canadian monetary policy shocks. By the end of the four-year horizon, Canadian shocks contribute 66.8% to fluctuations in the U.S. interest rate (R^*), 62.0% in the U.S. general price level (P^*), and 45.0% in U.S. output (y^*). Of these numbers, Canadian monetary policy shocks account for 17.3% in R^* , 15.2% in P^* , and 19.3% in y^* . Clearly, these results are at odds with the actual relationship between the Canadian economy and U.S. economy.

Figure 3 reports other anomalous results. The first column of graphics displays some dynamic responses to a shock originated from the production sector in the Canadian economy. As can be seen, this shock generates persistent negative response of the exchange rate (Exc) and substantial rise in both the Canadian interest rate (R) and U.S. interest rate (R^*). In fact, among all shocks this Canadian shock leads to the largest responses of both Exc and R^* . Yet it seems odd that large fluctuations in the U.S. interest rate (R^*) are caused mainly by a shock originated in the Canadian economy. Even for the Canadian exchange rate and interest rate, a priori belief is that its movements are caused mainly by fluctuations in the U.S. interest rate rather than by a Canadian shock.

These anomalous results are the artifacts of inappropriate treatment of the relationship between the Canadian economy and U.S. economy. Indeed, when Canada is taken seriously as a small economy and the exogeneity restriction is imposed, fluctuations in both the Canadian interest rate and exchange rate are caused largely by shocks originated in the U.S. economy, not

the Canadian economy. The second column of graphics in Figure 3 displays such results. These are the responses to a shock outside the Canadian economy. The fall in the Canadian dollar (Exc) and the rise in the Canadian interest rate (R) are associated with the rise in the U.S. interest rate (R*). Such a pattern, in contrast to the first column of Figure 3, one would normally expect: the Canadian dollar tends to fall when the U.S. interest rate increases, whereas the Canadian interest rate tends to follow fluctuations in the U.S. interest rate.

7. Conclusion

Finite-sample inference has become increasingly important in multivariate time-series models. While some restrictions on lagged relationships in structural VAR modeling are called for on economic grounds, obtaining exact finite-sample inference may be difficult.

This paper has developed block Monte Carlo methods to obtain ML estimates and Bayesian inference with less computational burden and has discussed the scope of their applicability. It is argued that a good many examples in the existing literature can be characterized by contemporaneously recursive blocks, and these blocks can be treated as independent for the purposes of estimation and inference. Meanwhile, concrete examples are used to show that results can be misleading if researchers fail to impose restrictions on lagged relationships that clearly make economic sense.

References

- Bernanke, Ben S., 1986. "Alternative Exploration of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy* 25, 49-99.
- Bernanke, Ben and Ilian Mihov, 1996. "Measuring Monetary Policy," *manuscript* (Princeton University).
- Bernanke, Ben S., Mark Gertler, and Mark Watson, 1997. "Systematic Monetary Policy and the Effects of Oil Price Shocks," *Brookings Papers on Economic Activity* 1, 91-142.
- Blanchard, Olivier J., 1989. "A Traditional Interpretation of Macroeconomic Fluctuations," *American Economic Review* 79 (5), 1147-1164.
- Blanchard, Olivier J. and Stanley Fischer, 1989. *Lectures on Macroeconomics*, the MIT Press, Cambridge, Massachusetts.
- Blanchard, Olivier J. and Mark Watson, 1986. "Are Business Cycles All Alike?" In: Gordon, Robert (Ed.), *The American Business Cycle: Continuity and Change*. University of Chicago Press, Chicago, 123-156.
- Canova, Fabio, 1991. "The Sources of Financial Crisis: Pre- and Post-Fed Evidence," *International Economic Review* 32, 689-713.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans, 1996. "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds," *Review of Economics and Statistics* 78 (1): 16-34.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans, 1997. "Monetary Policy Shocks: What Have We Learned and To What End?" forthcoming in the *Handbook of Macroeconomics* (Eds. John Taylor and Michael Woodford).
- Cochrane, John H., 1996. "What Do the VARs Mean? Measuring the Output Effects of Monetary Policy?" *manuscript* (University of Chicago).
- Cushman, David.O. and Tao Zha, 1997. "Identifying Monetary Policy in a Small Open Economy Under Flexible Exchange Rates," *Journal of Monetary Economics* 39 (3), 433-448.
- Dias, Francisco C., Jose A.F. Machado, and Maximiano R. Pinheiro, 1996. "Structural VAR Estimation with Exogeneity Restrictions," *Oxford Bulletin of Economics and Statistics* 58, 417-22.
- Doan, T.A., 1992. *RATS User's Manual Version 4*, Estima.
- Eichenbaum, Martin, and Charles Evans, 1995. "Some Empirical Evidence on the Effects of

- Shocks to Monetary Policy on Exchange Rates,” *Quarterly Journal of Economics* 110, 975-1009.
- Fisher, Franklin M., 1966. *The Identification Problem in Econometrics*, McGraw-Hill, Inc., New York, New York.
- Gali, J., 1992. “How Well Does the IS-LM Model Fit Postwar U.S. Data?” *Quarterly Journal of Economics*, May, 709-738.
- Geisser, S., 1965. “A Bayes Approach for Combining Correlated Estimates,” *Journal of American Statistical Association* 60, 602-607.
- Genberg, H., M.K. Salemi and A. Swoboda, 1987. “The Relative Importance of Foreign and Domestic Disturbances for Aggregate Fluctuations in the Open Economy,” *Journal of Monetary Economics* 19, 45-67.
- Geweke, John, 1989. “Bayesian Inference in Econometric Models using Monte Carlo Integration,” *Econometrica* 58, 1317-1339.
- Gordon, David B. and Eric M. Leeper, 1994. “The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification,” *Journal of Political Economy* 102, 1228-1247.
- Kilian, Lutz, 1997. “Small-Sample Confidence Intervals for Impulse Response Function,” forthcoming, *Review of Economics and Statistics*.
- Ingram, Beth F. and Charles H. Whiteman, 1994. “Supplanting the “Minnesota” Prior Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors,” *Journal of Monetary Economics* 34, 497-510.
- Leamer, E. E., 1978. *Specification Searches*, John Wiley & Sons, Inc.
- Leeper, Eric M., 1997. “Narrative and VAR Approaches to Monetary Policy: Common Identification Problems,” forthcoming, *Journal of Monetary Economics*.
- Leeper, Eric M. and David B. Gordon, 1992. “In Search of the Liquidity Effect,” *Journal of Monetary Economics* 29, 341-369.
- Leeper, Eric M. and Christopher A. Sims, 1994. “Toward a Modern Macroeconomic Model Usable for Policy Analysis,” *NBER Macroeconomics Annual*, 81-117.
- Leeper, Eric M., Christopher A. Sims, and Tao Zha, 1996. “What Does Monetary Policy Do?” *Brookings Papers on Economic Activity* 2, 1-63.
- Levy, D., S. Dutta and M. Bergen, 1996. “Shock Persistence and Price Rigidity,” *manuscript* (University of Chicago).

- Litterman, Robert B., 1986. "Forecasting With Bayesian Vector Autoregressions – Five Years of Experience," *Journal of Business and Economic Statistics* 4, 25-38.
- McCallum, Bennett T., 1989. *Monetary Economics: Theory and Policy*, Macmillan, New York, New York.
- Pagan, A.R., 1993. "Three Econometric Methodologies: An Update," in *Surveys in Econometrics* ed. by L. Oxley, D.A.R. George, C.J. Roberts and S. Sayer, 30-41.
- Pagan, A. and J. Robertson, 1995. "Resolving the Liquidity Effect," *Federal Reserve Bank of St. Louis Review* 77 (3), 33-54.
- Racette, D. and J. Raynauld, 1992. "Canadian Monetary Policy: Will the Checklist Approach Ever Get Us to Price Stability," *Canadian Journal of Economics* 25, November, 819-838.
- Sims, Christopher A., 1980. "Macroeconomics and Reality," *Econometrica* 48, 1-48.
- _____, 1986. "Are Forecasting Models Usable for Policy Analysis," *Quarterly Review of the Federal Reserve Bank of Minneapolis*, Winter.
- Sims, Christopher A. and Tao Zha, 1995. "Does Monetary Policy Generate Recessions?" *manuscript* (Yale University and Federal Reserve Bank of Atlanta).
- _____, 1996. "Bayesian Methods in Dynamic Multivariate Models," forthcoming, *International Economic Review*.
- _____, 1997. "Error Bands for Impulse Responses," *manuscript* (Yale University and Federal Reserve Bank of Atlanta).
- Strongin, Steven, 1995. "The Identification of Monetary Policy Disturbances: Explaining the Liquidity Puzzle," *Journal of Monetary Economics* 34 (3), 463-497.
- Uhlig, Harald, 1997. "What are the Effects of Monetary Policy? Results From an Agnostic Identification Procedure," *manuscript* (Tilburg University).
- Waggoner, Daniel F. and Tao Zha, 1997. "Normalization, Probability Distribution, and Impulse Responses," *Federal Reserve Bank of Atlanta Working Paper 97-11*.
- Zellner, Arnold, 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias," *Journal of the American Statistical Association* 57, 348-368.
- _____, 1971. *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, Inc. New York.
- _____, 1985. "Bayesian Econometrics," *Econometrica* 53, 253-269.

Zellner, Arnold and Franz Palm, 1974. "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics* 2, 17-54.

Table 1. Decomposition of Forecast Variances Attributed to All Canadian Shocks*

Months	12	24	48
R*	35.2 (6.8)	40.9 (12.7)	66.8 (17.5)
P*	17.4 (1.2)	30.7 (3.6)	62.0 (15.2)
Y*	36.2 (18.2)	46.5 (19.3)	45.0 (19.3)

*Numbers in parentheses are percentages of the decompositions that are attributed to Canadian monetary policy shocks.

Figure 1. Dynamic responses to a Poil shock.

Figure 2. Dynamic responses to a Canadian monetary policy shock.

Figure 3. Dynamic responses to shocks emanating from the Canadian and U.S. economies. The first column of graphics corresponds to the model with no exogeneity restriction, and the second column corresponds to the model with the exogeneity restriction.