Identification, Vector Autoregression, and Block Recursion

Tao Zha

Federal Reserve Bank of Atlanta
Working Paper 96-8
August 1996

Abstract: In the applications of identified VAR models, finite-sample properties are not obvious to obtain when identifying restrictions are imposed on some lagged relationships. As a result, researchers have either left lagged relationships unrestricted even though some restrictions clearly make economic sense or failed to provide correct inference of the estimates. We extend the Bayesian methodology in the existing literature to these cases and develop the blockwise Monte Carlo methods. We show how to implement these methods to obtain the estimation and inference.

JEL classification: C11, C15, C32, C50

Key words: Contemporaneous recursive blocks; identifying restrictions; likelihood; finite samples, posterior; blockwise Monte Carlo methods

---

The author has benefited from discussions with and suggestions by numerous people, especially Roberto Chang, Frank Diebold, Jon Faust, Eric Leeper, Adrian Pagan, Chris Sims, and Chuck Whiteman. He also thanks David Petersen for research assistance. The views expressed here are those of the author and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the author's responsibility.

Please address questions of substance to Tao Zha, Research Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713, 404/521-8353, 404/521-8956 (fax), tzha@frbatlanta.org.

Questions regarding subscriptions to the Federal Reserve Bank of Atlanta working paper series should be addressed to the Public Affairs Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713, 404/521-8020, http://www.frbatlanta.org.

1. Introduction

Identification, the ability to distinguish the behavior of one agent from other agents' behaviors in the actual economy, has been an important issue in the empirical literature. Many multivariate dynamic models, the vector autoregressive (VAR) models in particular, have attempted to achieve an identification while avoiding the "incredible" assumptions cautioned by Sims [1980]. Consequently, the restrictions imposed in these models have focused on contemporaneous relationships between variables or between residuals in a system of equations (Gordon and Leeper [1994], Pagan and Robertson [1994]).

Contemporaneous relationships deserve special attention for several reasons. First, many of these relationships can be justified either by traditional static frameworks (King, Plosser, Stock and Watson [1991], Gali [1992]) or by dynamic stochastic general equilibrium models (Sims and Zha [1996]). Second, without imposing other restrictions on the coefficients of lagged variables, one avoids potentially unreasonable assumptions and allows the data to reveal the dynamic patterns of different variables. Third, in a class of VAR models, leaving the coefficients of lagged variables unrestricted simplifies a statistical procedure considerably in the estimation and inference of structural parameters. This simplification is obtained because the likelihood as a function of contemporaneous structural parameters depends on the data only through the estimated covariance matrix of reduced-form residuals. Furthermore, given the contemporaneous structural parameters, the rest of structural parameters are simply a linear transformation of the reduced-from parameters (Sims and Zha [1995])[1].

These convenient statistical properties break down as we impose restrictions on both contemporaneous and lagged relationships. Yet in many empirical applications, restrictions on some coefficients of lagged variables are not unreasonable; on the contrary, they are necessary precisely on the grounds of economic reasoning. Failing to impose these restrictions for the convenience of statistical inference may result in misleading conclusions. For example, one crucial restriction in international macroeconomics is that changes in the variables of a small open economy exert little impact on the "rest of the world" (Genberg, Salemi and Swoboda [1987], Papell [1989], Dungey and Pagan [1996]). As shown in Cushman and Zha [1995], without imposing this restriction in empirical models, one would get the "exchange rate puzzle" in the case of Canada — the contractionary domestic monetary policy shocks depreciate the home currency.

On the other hand, some previous works have used approximate procedures for statistical inference when restrictions on certain lagged relationships are present. Some of these procedures lack clear theoretical justifications, especially in small samples. One example is the method used in Leeper and Gordon [1992] when money stock is treated exogenous in their VAR model. They use the standard procedure of Doan [1992] to generate the error bands of impulse responses without first taking explicit account of exogenous restrictions on money stock.[2]

---

[1] Such a simple mapping between reduced-form parameters and structural parameters can be preserved in some other kinds of restrictions such as the long-run restrictions in VAR models (Blanchard and Quah [1989], Faust and Leeper [1995]). But in general, as shown in this paper, the mapping is not nearly as simple when identifying restrictions go beyond contemporaneous relationships.

[2] The author wishes to thank Eric Leeper for bringing out this point.

In the existing applied works, there are numerous examples of requiring restrictions on some lagged relationships in certain equations within the whole system. In the literature of macroeconomics and industrial organization, a typical assumption is that some disaggregated variables (such as an individual firm's balance sheets) do not enter as explanatory variables the equations that describe the aggregate behavior (such as the sales in the whole industry) (Pagan [1993], Gilchrist and Zakrajsek [1995], Levy, Dutta and Bergen [1995]). In the applications of financial economics, the "world interest rate" serves as a driving force in the term structure of a particular country, but changes in the yields in that country are seldom considered to be influential on the determination of the world interest rate (Pagan [1995], Pagan, Hall and Martin [1995]).

This paper develops finite-sample methods that can be readily applied to the aforementioned and other examples. The methods extend the general Bayesian framework of Sims and Zha [1995] to situations wherein an original model is composed of blocks that are recursive in the coefficients of *contemporaneous* variables and wherein the lag structure, due to the possible restrictions on lagged behaviors, may change from block to block. We show that although the convenient Bayesian procedures derived by Sims and Zha [1995] cannot be used for estimation and inference of a whole system when the lag structure differs across blocks, it can be modified for and applied to each block independently when the blocks are contemporaneously recursive.[3] Our

[3] The advantage of Bayesian approaches in the presence of possible nonstationarity has been thoroughly discussed in the literature (see, for example, Sims [1988], DeJong and Whiteman [1991], Sims and Uhlig [1991], and Phillips [1994]). Our approach follows likelihood principle. To inform the reader of the likelihood, we use a non-informative prior; such a prior serves as a reference prior so that one can quantify one's real prior knowledge in particular problems as in Litterman [1986] and Ingram and Whiteman [1994].

methods work even when the coefficients of lagged variables in the original system have no block recursive relationships.

To implement these Bayesian methods, we construct algorithms that allow one to obtain finite-sample inference through Monte Carlo simulations.[4] We undertake a couple of concrete computational exercises to demonstrate how feasible it is to use our algorithms when blocks have different lag structures. Meanwhile, we document the enormous gain in computation when an individual block of equations can be estimated independent of other blocks.

After laying out the general framework in the next section, we begin, in Section 3, with the recursive structure that has been commonly used in previous empirical works and show how the Bayesian procedure of Sims and Zha [1995] is extended to the blockwise estimation and inference of individual blocks of equations. Section 4 uses a concrete example to illustrate the implementation of the method of Section 3. In Section 5, this method is modified and generalized for a wider range of economic problems. In Section 6, we apply the generalized method to the identification of monetary policy during 1979-82.

2. Setup of General Model

We consider a general model (ignoring the predetermined variables in our notations):

$$A(L)y(t) = \varepsilon(t), \quad t = 1, \ldots, T, \tag{1}$$

[4] Kilian [1996] performs various versions of classical bootstrap method to construct confidence intervals of impulse responses in finite samples. For extensive discussions of philosophical issues concerning the differences between classical and Bayesian approaches in finite samples, see Sims and Zha [1995].

where $A(L)$ is an $M \times M$ matrix of polynomials in lag operators, $y(t)$ an $M \times 1$ vector of observed variables, $\varepsilon(t)$ an $M \times 1$ vector of structural disturbances, and $A(0)$, the coefficient matrix of L^0 in $A(L)$, is nonsingular. We assume $\varepsilon(t)$ is Gaussian with

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

where I is the identity matrix with dimension M. Let us 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

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

Accordingly we divide system (1) into a set of blocks:

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

where $A_i(L)$ is the matrix $(A_{i1}(L),\dots,A_{in}(L))$ and $\varepsilon_i(t)$ is an $m_i \times 1$ vector of corresponding disturbances in block i. While maintaining an assumption that the lag structure is the same across the equations within a block, we allow possible identifying restrictions on the lagged coefficients in $A(L)$ so that the lag structure can be different across blocks.

3. Strongly Contemporaneous Recursive Blocks

We begin our analysis with the model that has strongly contemporaneous recursive blocks.

Definition 1. Model (1) has strongly contemporaneous recursive blocks or $A(L)$ is strongly contemporaneous block recursive if $A_{ij}(0) = 0$ for $i > j$ and $A_{ij}(0)$ is unrestricted for $j > i$.

Denoting the block diagonal coefficient matrix of L^0 in $A(L)$ by

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

and rewriting model (1) as

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

we divide (4) into a set of normalized blocks, each normalized block corresponding to the original block in (3):

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

with

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

where $\mathbf{0}_{i-}$ the matrix of zeros with dimension $m_i \times (m_1 + \ldots + m_{i-1})$, I_i the identity matrix with dimension m_i, $\mathbf{0}_{i+}$ the matrix of zeros with dimension $m_i \times (m_{i+1} + \ldots + m_m)$, and $y_i(t)$ an $m_i \times 1$ vector of observed contemporaneous variables in block i.

System (5) comprises n blocks of equations. In each block i, the right hand side of the equations contains the contemporaneous variables $y_j(t)$ only for $j > i$. And the normalized disturbances $v(t)$ have the following block characteristic:

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

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

Framework (5) is widely used in the existing literature. The leading case is a class of VAR models with a Choleski decomposition of the estimated covariance matrix of residuals. In these cases, each equation can form a block. A more general use of (5), however, concentrates on cases in which the contemporaneous coefficient matrix $A_{ii}(0)$ is non-recursive or the lag structure is different across the blocks or both.

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

$$\underset{T\times m_i}{\mathbf{Y}_i} = \underset{T\times k_i}{\mathbf{X}_i}\ \underset{k_i\times m_i}{\mathbf{C}_i} + \underset{T\times m_i}{\mathbf{v}_i}\ ,\quad i=1,\dots,n, \tag{10}$$

where $\mathbf{Y}_i$ is a matrix of observations of contemporaneous variables, $\mathbf{X}_i$ a matrix of observations of lagged variables plus contemporaneous variables from other blocks ($\mathbf{Y}_j$'s $(j>i)$) plus deterministic variables (which are ignored in our notations), and $\mathbf{C}_i$ the coefficient matrix corresponding to $\mathbf{X}_i$. It can be seen from (1) that the conditional p.d.f. of $y(t)$ is

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

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

$$|A_d(0)|^T \exp\left[-\frac{1}{2}\sum_t (A(L)y(t))'(A(L)y(t))\right]$$

$$\propto \prod_i |A_{ii}(0)|^T \exp\left[-\frac{1}{2}\operatorname{trace}(S_i(\mathbf{C}_i)A_{ii}(0)'A_{ii}(0))\right]$$

$$\propto \prod_i |A_{ii}(0)|^T \exp\left[-\frac{1}{2}\operatorname{trace}\left(S_i(\hat{\mathbf{C}}_i)A_{ii}(0)'A_{ii}(0) + \left(\mathbf{C}_i-\hat{\mathbf{C}}_i\right)'\mathbf{X}_i'\mathbf{X}_i\left(\mathbf{C}_i-\hat{\mathbf{C}}_i\right)A_{ii}(0)'A_{ii}(0)\right)\right], \tag{11}$$

where

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

In the first line of expression (11), we have used the equality $|A_d(0)| = |A(0)|$ because $A(0)$ is block recursive. The second line of (11) indicates that this equality is crucial for dividing the likelihood for the whole system into separate likelihood functions for individual blocks. From the third line of (11), it is clear that the concentrated likelihood for $A_{ii}(0)$ is:

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

With the likelihood functions (11) and (13), let us turn to the Bayesian estimation and inference. To inform the reader of overall likelihood shape, we consider only a diffuse prior on $(A_{ii}(0), \mathbf{C}_i)$. To eliminate the possible discrepancies between posterior modes and maximum likelihood (ML) estimates, we follow Sims and Zha [1995] and take up the improper prior $|A_{ii}(0)|^{k_i}$ for the elements of $A_{ii}(0)$. The following two propositions and algorithm establish our ***blockwise*** method.

Proposition 1. The Bayesian posterior mode of $(A_{ii}(0), \mathbf{C}_i)$ is exactly the maximum likelihood estimate $\left(\hat{A}_{ii}(0), \hat{\mathbf{C}}_i\right)$, where $\hat{A}_{ii}(0)$ maximizes (13) and $\hat{\mathbf{C}}_i$ is simply the OLS estimate expressed in (12).

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

$$|A_{ii}(0)|^{T+k_i} \exp\left[-\frac{1}{2}\mathrm{trace}\left(S_i(\hat{\mathbf{C}}_i)A_{ii}(0)'A_{ii}(0) + \left(\mathbf{C}_i - \hat{\mathbf{C}}_i\right)'\mathbf{X}_i'\mathbf{X}_i\left(\mathbf{C}_i - \hat{\mathbf{C}}_i\right)A_{ii}(0)'A_{ii}(0)\right)\right]. \tag{14}$$

One can easily derive the marginal posterior p.d.f. of $A_{ii}(0)$ from (14), which is the same as the concentrated likelihood (13). Therefore, the posterior mode and the ML estimate of $A_{ii}(0)$ coincide.

It is also clear from (11) that the posterior distribution of vec($\mathbf{C}_i$) (the vectorized column of $\mathbf{C}_i$) conditional on $A_{ii}(0)$ is Gaussian with mean vec($\hat{\mathbf{C}}_i$) and covariance matrix

$$\left(A_{ii}(0)'A_{ii}(0)\right)^{-1} \otimes \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}.$$

Hence the posterior mode of $\mathbf{C}_i$ is just $\hat{\mathbf{C}}_i$.

Q.E.D.

Proposition 2. Given $A_{ii}(0)$ and $\mathbf{C}_i$ or $C_i(L)$, the structural parameter matrix $A(L)$ in model (1) can be calculated block by block according to

$$A_i(L) = A_{ii}(0)\left(\left(\mathbf{0}_{i-}, I_i, \mathbf{0}_{i+}\right) - C_i(L)\right), \quad i = 1, \ldots, n. \tag{15}$$

Proof. Equation (15) follows immediately from (6).

Q.E.D.

Given the posterior distribution of $A_{ii}(0)$ and $C_i(L)$ or $\mathbf{C}_i$ for $i = 1, \ldots, n$, the finite-sample inference of model parameters $A(L)$ or a function of $A(L)$ (denoted by $f(A(L))$ can be obtained by generating Monte Carlo (MC) samples block by block. We summarize the algorithm, following, of generating these MC samples.

Algorithm 1. The computational procedure involves several steps:

(a) generate samples of $A_{ii}(0)$ by drawing from the marginal posterior p.d.f. (13);

(b) conditional on drawn $A_{ii}(0)$'s, sample $\mathbf{C}_i$ or $C_i(L)$ from its Gaussian distribution;

(c) for the given MC samples of $A_{ii}(0)$ and $C_i(L)$, calculate samples of $A_i(L)$ by (15) in Proposition 2;

(d) calculate $f(A(L))$'s from the 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 crucial steps in Algorithm 1 are (a)-(c), and they can be all done block by block. The only rub with this algorithm in practice is step (a). Since (13) is in general not of any standard distribution, direct draws from it become infeasible. To overcome this difficulty, let us first consider situations in which there is a one-one mapping between $A_{ii}(0)$ and Σ_{ii} in (9). In such cases, we transform $A_{ii}(0)$ into Σ_{ii} in likelihood (11) according to (9). Using the improper prior $|\Sigma_{ii}|^{-\frac{k_i}{2}}$ for Σ_{ii}, we obtain the marginal posterior p.d.f. of Σ_{ii} as[5]

$$|\Sigma_{ii}^{-1}|^{\frac{T}{2}} \exp\left[-\frac{1}{2}\mathrm{trace}\left(S_i\left(\hat{\mathbf{C}}_i\right)\Sigma_{ii}^{-1}\right)\right]. \tag{16}$$

From (9) and (11), one can easily see that (16) is also the concentrated likelihood function for Σ_{ii}.

The p.d.f. (16) has an inverted Wishart form and Σ_{ii}^{-1} has the following Wishart distribution[6]

$$\mathit{Wishart}\left(S_i^{-1}(\hat{\mathbf{C}}_i), T - m_i - 1, m_i\right). \tag{17}$$

Monte Carlo samples of $A_{ii}(0)$ can be generated by first drawing Σ_{ii}^{-1} from (17) and then transforming the drawn Σ_{ii}^{-1} back into $A_{ii}(0)$ according to (9). Bayesian inference of $f(A(L))$ is then

---

[5] The prior $|\Sigma_{ii}|^{-\frac{k_i}{2}}$ for Σ_{ii} is implied by the prior $|A_{ii}(0)|^{k_i}$ for $A_{ii}(0)$. It is a "flat" Wishart p.d.f. in the sense of Geisser [1965] and in a univariate case is simply an ignorance prior discussed in Leamer [1978] (pp. 78-84).

[6] See, for example, pp. 389-396 in Zellner [1971].

computed by following steps (b)-(e) in Algorithm 1.

When the mapping between $A_{ii}(0)$ and Σ_{ii} does not possess a one-one relation, it makes no sense to draw Σ_{ii}^{-1} from (17) because $A_{ii}(0)$ can no longer recover uniquely from Σ_{ii}. We can, however, apply the weighted Monte Carlo method of Sims and Zha [1995] to each of these blocks (i.e., blocks that have no one-one relation between $A_{ii}(0)$ and Σ_{ii}). Specifically, we draw $A_{ii}(0)$ from the asymptotic Gaussian distribution approximated by the second-order Taylor expansion of the logarithm of (13) at its mode and then repeat steps (b)-(d) in Algorithm 1. Prior to proceeding with step (e) in Algorithm 1, we weight all the drawn samples of $f(A(L))$ by the ratio of a product of (13)'s for these blocks to a product of those approximate Gaussian p.d.f.'s.[7]

4. Money-Income Relationship Revisited

There exists a long list of applications that fall into the class of models discussed in the previous section. The models studied by Genberg, Salemi and Swoboda [1987], Racette and Raynauld [1992], Leeper and Gordon [1992], and Levy, Dutta and Bergen [1995] are a few examples. All these models have strongly contemporaneous recursive blocks and share the feature that the variables in some blocks do not enter the equations in other blocks. Although Genberg, Salemi and Swoboda [1987] and Racette and Raynauld [1992] compute the impulse responses from the estimates of their model parameters, they do not provide any error bands for the computed

[7] 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}. Especially, when such a nonlinear mapping is complicated and requires some non-trivial computing time to transform Σ_{ii} back into $A_{ii}(0)$, the weighted procedure may be efficient as compared to the procedure with direct draws from Wishart.

responses. Levy, Dutta and Bergen [1995] report asymptotic confidence bands for their estimated impulse response functions. But as Kilian [1996] shows, asymptotic methods can perform very poorly by classical criteria in finite samples.

Since impulse responses are simply a special function of model parameters with the functional form $f(A(L)) = A^{-1}(L)$, the blockwise method of Section 3 can be readily used for computing the error bands in all these models. For the sole purpose of illustrating how our method can be implemented, let us turn to a simple, bivariate model of money-income relationship. The principles of implementation are applicable in more complex models such as those discussed in the previous paragraph.

The bivariate model we use is based on Christiano and Ljungqvist [1988] with their sample period 1948:1-1985:12. The monthly data are seasonally adjusted M1 (M) and the index of seasonally adjusted industrial production (y), both in logarithm. We take up the assumption that money "M" is exogenous to output "y" and has predictive power in "y".[8] This implies that contemporaneous and lagged M's enter the y equation (which we let be the first block in this bivariate system) but the M equation (the second block) includes none of y's. Thus the system has strongly

---

[8] This assumption is consistent with the findings in the existing literature (Christiano and Ljungqvist [1988], Stock and Watson [1989], Friedman and Kuttner [1993], Thoma [1994]). In fact, it can be easily verified that in the y-M model with six lags and Choleski decomposition and no other restrictions, the impulse responses of "M" to a shock in "y" are insignificant (gauged by the .95 probability bands) for almost the entire time horizon of, say, 8 years, while the responses of "y" to a shock in "M" are highly significant. This result holds no matter how the Choleski decomposition is ordered. It is worth noting that while showing to the reader this concrete example of how to implement our method, we do not take any position regarding the appropriateness of modeling money-income relationships in the bivariate framework used in the previous works.

contemporaneous recursive blocks and the lag structure of the M block is different from that of the y block. Apparently, the procedures described in Doan [1992] or in Sims and Zha [1995] cannot be put to direct use. To see how to use the blockwise method of Section 3, we write out explicitly the normalized version of this bivariate system according to (5):

$$y(t) = \delta M_t + \sum_{s=1}^{6} \theta_s y(t-s) + \sum_{s=1}^{6} \mu_s M(t-s) + c_y + v_1(t), \tag{18}$$

$$M(t) = \sum_{s=1}^{6} \gamma_s M(t-s) + c_M + v_2(t), \tag{19}$$

where c_y and c_M are the constant coefficients. Using the notations in Section 3, we have that Σ_{11} is the variance of v_1, Σ_{22} the variance of v_2, and

$$\mathbf{C}_1 = \left(\delta, \theta_1, \ldots, \theta_6, \mu_1, \ldots, \mu_6, c_y\right), \;\; \mathbf{C}_2 = \left(\gamma_1, \ldots, \gamma_6, c_M\right).$$

The posterior mode of $\mathbf{C}_i$ $(i = 1,2)$ is just the ML estimate $\hat{\mathbf{C}}_i$ $(i = 1,2)$ given in (12). The posterior mode of $A_{ii}(0)$ $(i = 1,2)$, according to (13), is $\Big/\sqrt{S_i\left(\hat{\mathbf{C}}_i\right)}$ $(i = 1,2)$. By Proposition 2, one can derive the posterior mode or ML estimate $\hat{A}(L)$ and compute the estimated impulse response functions as $\hat{A}^{-1}(L)$. Figure 1 plots these estimated responses (solid lines) for the horizon of 48 months.

To compute the probability bands for impulse response functions, one could first draw $A_{ii}(0)$ $(i = 1,2)$ from its approximate Gaussian distribution, calculate the weights, and then use the weighted procedure of Section 3. But the one-one mapping between $A_{ii}(0)$ and Σ_{ii} is so convenient that we instead draw each Σ_{ii}^{-1} $(i = 1,2)$ directly from one-dimensional Wishart distribution

(17) (or the general Chi-square distribution) and calculate the corresponding $A_{ii}(0)$ by

$$A_{ii}(0)=\sqrt{\Sigma_{ii}^{-1}},\ i=1,2.$$

Steps (b)-(e) in Algorithm 1 are then repeated for each draw. From the 5,000 MC samples we compute the equal-tailed Bayesian posterior .68 and .95 probability bands for impulse responses. The dotted lines in Figure 1 display the results; the narrower band contains .68 probability and the wider one .95. Note that the small panel on the left hand bottom in Figure 1 is the outcome of our restriction that movements in "y" have no influence on "M" in equation (19).

5. Weakly Contemporaneous Recursive Blocks

In the previous sections, we consider cases in which all the elements of submatrix $A_{ij}(0)$ $(j>i)$ are unrestricted. There are three compelling reasons for doing so. First, such cases have been widely used in the existing literature; second, even in these simple cases, the finite-sample inference is either not provided or incorrectly computed in previous works; third, the blockwise method of Section 3 provides the basic approach that proves useful for other cases.

In general, models that attempt to identify certain behaviors in the actual economy are very likely to go beyond the feature of strongly contemporaneous recursive blocks. In the recent literature of policy analysis (Gordon and Leeper [1994], Cushman and Zha [1995], Bernanke and Mihov [1996], Dungey and Pagan [1996]), for example, some restrictions on $A_{ij}(0)$ $(j>i)$ are required in order to achieve an identification of the policy reaction behavior. Such restrictions make it impossible to rely on Proposition 1 and Algorithm 1 if we are to derive the blockwise estimation and inference. This is simply because $\hat{C}_i(L)$ or $\hat{\mathbf{C}}_i$ in (12) ignores the restrictions on the elements in $A_{ij}(0)$ for $j > i$ and thus is no longer the ML estimate of $C_i(L)$. To take account of these restrictions, we follow the basic idea of Section 3 and establish generalized blockwise Monte Carlo methods that are applicable to both types of models described in Definitions 1 and 2. To begin our analysis, let us have the following definition.

Definition 2. Model (1) has weakly contemporaneous recursive blocks or $A(L)$ is weakly contemporaneous block recursive if $A_{ij}(0) = 0$ for $i > j$ and if there are further restrictions on some elements in $A_{ij}(0)$ for $j > i$.

Now, define

$$D(L) = A_d^{-1}(0)A(L)\ ,\ \ F(L) = -\big(D(L) - D(0)\big).$$

In agreement with the divisions in (3), we partition $D(L)$ into a column vector of n block submatrices $D_i(L)$'s $(i = 1,\ldots,n)$ and similarly $F(L)$ into a column vector of $F_i(L)$'s $(i = 1,\ldots,n)$ so that

$$D_i(L) = A_{ii}^{-1}(0)A_i(L), \tag{20}$$

$$F_i(L) = -\big(D_i(L) - D_i(0)\big). \tag{21}$$

Thus, (5) can be rearranged as

$$D_i(0)y(t) = F_i(L)y(t) + v_i(t),\ \ i = 1,\ldots,n,\ \text{ all } t\ . \tag{22}$$

To write (22) in compact matrix form, we let $\mathbf{Z}_i$ be a $T \times m_i$ matrix of the left hand side variables of (22), $\mathbf{W}_i$ be a $T \times q_i$ matrix of observations on the right hand side of (22) where

$$q_i = k_i - m_{i+1} - \ldots - m_n\ ,$$

and $\mathbf{F}_i$ be the $q_i \times m_i$ matrix form of $F_i(L)$ corresponding to $\mathbf{W}_i$. We thus have the matrix version of (22):

$$\mathbf{Z}_i = \mathbf{W}_i\mathbf{F}_i + \mathbf{v}_i . \tag{23}$$

Note that $\mathbf{Z}_i$ depends on both the observations $y(t)$ and the parameters in $D_i(0)$ or $A_{ij}(0)$ for $j > i$.

The following two theorems and algorithm establish our generalized blockwise Monte Carlo methods.

Theorem 1. For $i = 1,\ldots,n$, define

$$\hat{\mathbf{F}}_i = \left(\mathbf{W}_i'\mathbf{W}_i\right)^{-1}\mathbf{W}_i'\mathbf{Z}_i , \tag{24}$$

$$V_i(\mathbf{F}_i) = (\mathbf{Z}_i - \mathbf{W}_i\mathbf{F}_i)'(\mathbf{Z}_i - \mathbf{W}_i\mathbf{F}_i) . \tag{25}$$

The ML estimates of the parameters in $A_i(0)$ can be obtained (independent of all other parameters within and outside the block) by finding the elements in $A_i(0)$ that maximize

$$|A_{ii}(0)|^T \exp\left[-\frac{1}{2}\mathrm{trace}\left(\mathrm{V}_i\left(\hat{\mathbf{F}}_i\right)A_{ii}(0)'A_{ii}(0)\right)\right] , \tag{26}$$

and the ML estimate of $\mathbf{F}_i$ is the value of $\hat{\mathbf{F}}_i$ at the ML estimates of the elements in $A_{ij}(0)$ for $j > i$.

Proof. To reach the conclusion, let us rearrange the algebraic expression of likelihood function (11) in a different form:

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

Since the parameters in $\mathbf{Z}_i$ are only $A_{ij}(0)$ for $j > i$ within block i, $\hat{\mathbf{F}}_i$ in (24) as a function of parameters depends only on $A_{ij}(0)$ for $j>i$. Given the ML estimate of $A_{ij}(0)$ for $j>i$, it must be true from (27) that the ML estimate of $\mathbf{F}_i$ is simply the value of $\hat{\mathbf{F}}_i$ with the parameters in $A_{ij}(0)$ $(j>i)$ replaced by their ML estimates. These features connote that the elements in $A_i(0)$ that maximize (26) also maximize (27).

Q.E.D.

Theorem 2. With the diffuse prior $|A_{ii}(0)|^{q_i}$ on $(A_i(0),\mathbf{F}_i)$, the posterior joint distribution of $(A_i(0),\mathbf{F}_i)$ is

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

where

$$p(vec(\mathbf{F}_i)|A_i(0)) \propto \mathrm{N}\left(vec(\hat{\mathbf{F}}_i),(A_{ii}(0)'A_{ii}(0))^{-1}\otimes\left(\mathbf{W}_i'\mathbf{W}_i\right)^{-1}\right), \tag{28}$$

$$p(A_i(0)) \propto (26).$$

Proof. With likelihood (27), the posterior p.d.f. of $(A_i(0),\mathbf{F}_i)$ is proportional to

$$|A_{ii}(0)|^{T+q_i}\exp\left[-\frac{1}{2}\mathrm{trace}\left(V_i(\hat{\mathbf{F}}_i)A_{ii}(0)'A_{ii}(0)+\left(\mathbf{F}_i-\hat{\mathbf{F}}_i\right)'\mathbf{W}_i'\mathbf{W}_i\left(\mathbf{F}_i-\hat{\mathbf{F}}_i\right)A_{ii}(0)'A_{ii}(0)\right)\right]. \tag{29}$$

It is clear from (29) that the posterior p.d.f. of $\mathbf{F}_i$ *conditional* on $A_i(0)$ is proportional to

$$|A_{ii}(0)|^{q_i}\exp\left[-\frac{1}{2}\mathrm{trace}\left(\left(\mathbf{F}_i-\hat{\mathbf{F}}_i\right)'\mathbf{W}_i'\mathbf{W}_i\left(\mathbf{F}_i-\hat{\mathbf{F}}_i\right)A_{ii}(0)'A_{ii}(0)\right)\right],\quad i=1,\ldots,n. \tag{30}$$

Since $V_i\left(\hat{\mathbf{F}}_i\right)$ is independent of $\mathbf{F}_i$ (see (24) and (25)), (30) implies (28). Finally, if we integrate out $\mathbf{F}_i$ in the posterior joint p.d.f. (29) according to (30), , we obtain the marginal posterior probability density function of $A_i(0)$ which is proportional to (26).

Q.E.D.

The above theorems lay out the theoretical foundations on which computation of the Bayesian finite-sample inference of $A(L)$ or $f(A(L))$ is based. Specifically, we have, similar to Algorithm 1 in Section 3,

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

(a) draw $A_i(0)$ from the asymptotic Gaussian approximated by the second-order Taylor expansion of the logarithm of (26) at its peak;

(b) conditional on the drawn samples of $A_i(0)$, generate samples of F_i or $F_i(L)$ from the Gaussian distribution (28);

(c) given the samples of $A_i(0)$, calculate samples of $D_i(0)$ as $A_{ii}^{-1}(0)A_i(0)$ from (20); next, given the samples of $D_i(0)$ as well as $F_i(L)$, calculate samples of $D_i(L)$ from (21); finally, with the samples of $D_i(L)$ and $A_i(0)$, calculate samples of $A_i(L)$ from (20) for $i = 1,\ldots,n$.

(d) calculate samples of $f(A(L))$ from $A(L)$'s, and then weight each sample of $f(A(L))$ by the ratio of a product of (26)'s (for $i = 1,\ldots,n$) to a product of all the approximate Gaussian p.d.f.'s ($i = 1,\ldots,n$);

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

As in Algorithm 1, the crucial steps (a)-(c) in Algorithm 2 require only blockwise computations. We note that the methods developed in Theorems 1 and 2 and Algorithm 2 are applicable to all the cases discussed in Section 3. To be precise, we have the following

Theorem 3. If block i is strongly contemporaneous recursive to other blocks, the generalized blockwise method of this section (Theorems 1 and 2 and Algorithm 2) attains the same estimation and inference of block parameters as does the blockwise method of Section 3 (Proposition 1 and Algorithm 1).

Proof. We denote

$$A_{i+}(0) \text{ with the dimension } m_i \times (m_{i+1}+\ldots+m_n)$$

as the submatrix of the last $m_{i+1},\ldots,m_n$ columns in $A_i(0)$ so that

$$A_i(0) = \left(A_{ii}(0), A_{i+}(0)\right).$$

The parameter matrix in Theorems 1 and 2 and Algorithm 2 is

$$\left(A_{ii}(0), A_{i+}(0), F_i(L)\right), \tag{31}$$

and the parameter matrix in Proposition 1 and Algorithm 1 is

$$\left(A_{ii}(0), C_i(L)\right). \tag{32}$$

Since the elements in $A_{i+}(0)$ are unrestricted by assumption, the parameter matrices (31) and (32) have the one-one mapping:

$$C_i(L) = -A_{ii}^{-1}(0)A_{i+}(0) + F_i(L). \tag{33}$$

We arrive at the relation (33) because it is not hard to see from (6), (20), and (21) that

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

To prove this theorem, it is sufficient to show that after we transform (31) into (32), posterior p.d.f. (29) of parameter matrix (31) is the same as posterior p.d.f. (14) of parameter matrix (32). First, we note that since (5) and (22) are simply the two different arrangements of the same block of equations, likelihood functions (11) and (27) are actually identical. Second, given $A_{ii}(0)$, the mapping (33) between $C_i(L)$ and $\left(A_{i+}(0), F_i(L)\right)$ is a simple linear transformation, and the Jacobian of transformation of $\left(A_{i+}(0), F_i(L)\right)$ into $C_i(L)$ is

$$\left|A_{ii}(0)\right|^{m_{i+1}+\ldots+m_n} \equiv \left|A_{ii}(0)\right|^{k_i - q_i}. \tag{34}$$

The implied posterior p.d.f. of parameter matrix (32) from the distribution of (31) is simply the posterior p.d.f. (29) multiplied by the Jocabian $\left|A_{ii}(0)\right|^{k_i - q_i}$. The result leads to the exact form of (14) since likelihood functions (11) and (27) are identical.

Q.E.D.

When $A(L)$ is strongly contemporaneous block recursive, the method of Section 3 is strictly preferred because the estimation and inference of parameters in $A_{ij}(0)$ $(j > i)$ do not involve the maximization problem as the method of this section does. In practice, however, model (1) often has a mix of strongly and weakly contemporaneous recursive blocks. In these situations, Theorem 3 proves useful: for strongly contemporaneous recursive blocks, use the method of Section 3; for weakly contemporaneous recursive blocks, use the method of this section.

We are now in a position to discuss briefly an application of our blockwise Monte Carlo methods in some recent works. In the small open economy model of Cushman and Zha [1995], there are two contemporaneous blocks — the first block describes the behavior of the Canadian economy and the second block contains only the U.S. variables.[9] The lag structure is different across these two blocks. While the first block is weakly contemporaneous recursive, the second block is strongly recursive. The estimation and inference of functions of their model parameters such as the impulse responses can be easily computed by our blockwise MC methods, even though the method of Sims and Zha [1995] is not applicable there.

The blockwise MC methods also clear the way for computational efficiency. Let us use the identified 7-variable model by Gordon and Leeper [1994] as an illustration. Since all the equations in their model have exactly the same lag structure, the procedures of Sims and Zha [1995] and the blockwise methods can be both used to obtain the estimation and inference of, say, impulse response functions. But the gain in computing time is enormous with the blockwise methods. For example, the Gordon and Leeper model can be divided into 6 contemporaneous recursive blocks, some of which are weakly recursive. The ML estimation of their model parameters when treating the system as just one block takes about 80-100 times longer (depending on the starting point in maximization) than when dividing the system into these 6 blocks. This message is important because the estimation and inference of many a potentially large VAR system, which appear to be computationally prohibitive, can be feasibly done when contemporaneous recursive blocks are utilized.

[9] Note that there is no unique way of dividing the system. Further divisions can be made within each of the two blocks of Cushman and Zha's model. In general, more blocks imply more efficiency gains in computation.

6. Monetary Policy: 1979-1982

The Federal Reserve's "new" operating procedure, which began on October 6, 1979 and ended on October 9, 1982, has been to date the subject of intense interest.[10] One important empirical issue that is yet to be resolved relates to the estimation and inference of quantitative effects of monetary policy during 1979-1982. We follow the recent identified VAR approach to see how our blockwise Monte Carlo methods can be actually applied to this problem,

Let us consider the model that has four variables with six lags: output "y" (industrial output), price "P" (consumer price index bar shelter), the interest rate "R" (the 3-month treasure bill rate), and money stock "M" (M1).[11] The data are monthly with the sample period 1979:10 - 1982:9 and all the variables are in logarithm except R and. In our identification (indicated in Table 1), we assume, as in Gordon and Leeper [1994], that output and the price level do not respond to the financial variables contemporaneously (see the equations in the production sector in Table 1). In the equation of money demand (MD), we take up the functional form M-P = y-aR, which derives from many monetary models.

The monetary policy reaction function (the money supply equation in Table 1) allows the

---

[10] We use the quote "new" here because it is still questionable whether the Fed really changed its behavior from its previous policy (Poole [1982], Bryant [1983]). But this is not the place to address such a controversial issue; the paper here concerns primarily the new econometric methods.

[11] We also use the series of total reserves and federal funds rates in place of M1 and T-bill rates, the responses to monetary policy shocks change little quantitatively. It is worth noting that since the period 1979-82 is very short, we will greatly reduce the precision of estimates if we give up much of the degree of freedom. For example, when we try to include the series of commodity prices, the convergence in the maximum likelihood estimation becomes problematic.

7. Conclusion

Although finite-sample properties have become increasingly important in many multivariate time-series models, the difficulty of obtaining these properties occurs when we have restrictions on some lagged relationships in VAR models. Such difficulty has prevented previous works from providing correct inference or any inference of the estimates.

We argue that a good many models in the existing literature can be characterized by contemporaneous recursive blocks. Such a characteristic enables us to develop the blockwise Monte Carlo methods. We discuss the scope of their applicability and the potential of their use. We show how to implement these Bayesian methods in some actual economic problems and provide the practical algorithms for executing computations. Meanwhile, our methods point to a way of attaining the estimation and inference of potentially large models and can be modified for the VAR models with informative priors.

REFERENCES

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

Bernanke, Ben and Ilian Mihov, 1996. "Measuring Monetary Policy," *manuscript* (Princeton University).

Bryant, Ralph C., 1983. *Controlling Money*, The Brookings Institution, Washington D.C.

Christiano, L.J. and L. Ljungqvist, 1988. "Money Does Granger Cause Output in the Bivariate Money-Output Relation," *Journal of Monetary Economics 22*, 217-235.

Cook, Timothy, 1989. "Determinants of the federal funds rate: 1979-1982," *Economic Review*, Federal Reserve Bank of Richmond, January, 3-19.

Cushman, D.O. and T. Zha, 1995. "Identifying Monetary Policy in a Small Open Economy Under Flexible Exchange Rates," *Federal Reserve Bank of Atlanta Working Paper 95-7*.

DeJong, D.N. and C.H.Whiteman, 1991. "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series'," *Journal of Monetary Economics 28*, 221-254.

Doan, T.A., 1992. *RATS User's Manual Version 4*, Estima.

Dungey, Mardi and Adrian Pagan, 1996. "Towards a Structural VAR Model of the Australian Economy," *processed manuscript* (Australian National University).

Faust, J. and E.M. Leeper, 1995. "When Do Long-Run Identifying Restrictions Give Reliable Results?" to appear in *Journal of Business & Economic Statistics*.

Friedman, B.M. and K.N. Kuttner, 1993. "Another Look at the Evidence on Money-Income Causality," *Journal of Econometrics 57*, 189-203.

Gali, J., 1992. "How Well Does the IS-LM Model Fit Postwar U.S. Data?" *Quarterly Journal of Economics*, May, 709-738.

Gavin, William T. and Finn E. Kydland, 1996. "Endogenous Money Supply and the Business Cycle," *manuscript* (Federal Reserve Bank of St. Louis).

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.

Gilchrist, S. and E. Zakrajsek, 1995. "Investment and the Present Value of Profits: An Analysis Using Firm-Level Data," *manuscript* (Boston University).

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, 1996. "Small-Sample Confidence Intervals for Impulse Response Function," to appear in *Review of Economics and Statistics*.

King, Robert G., Charles I. Plosser, James H. Stock and Mark W. Watson, 1991. "Stochastic Trends and Economic Fluctuations," *American Economic Review 81*, 819-840.

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. and David B. Gordon, 1994. "In Search of the Liquidity Effect," *Journal of Monetary Economics 29*, 341-369.

Levy, D., S. Dutta and M. Bergen, 1995. "Shock Persistence and Price Rigidity," *manuscript* (University of Chicago).

Litterman, Robert B., 1986. "Forecasting with Bayesian Vector Autoregression — Five Years of Experience," *Journal of Business & Economic Statistics 4*, 25-38.

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.

________, 1995. "The Econometrics of Financial Markets," *processed manuscript*.

Pagan, A.R., A.D. Hall and V. Martin, 1995. "Modeling the Term Structure," *manuscript* (Australian National University).

Pagan, A. and J. Robertson, 1994. "Resolving the Liquidity Effect," to appear in *Federal Reserve Bank of St. Louis Review*.

Papell, D.H., 1989. "Monetary Policy in the United States Under Flexible Exchange Rates," *American Economic Review 79*, 1106-1116.

Phillips, P.C.B., 1994. "Nonstationarity Time Series and Cointegration: Recent Books and Themes for the Future," *manuscript* (Yale University).

Poole, William, 1982. "Federal reserve operating procedures, " *Journal of Money, Credit, and Banking 14*, 575-596.

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, C.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.

________, 1988. "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control 12*: 436-474.

Sims, C.A. and H. Uhlig, 1991. "Understanding Unit Rooters: A Helicopter Tour," *Econometrica 59*, 1591-1599.

Sims, C.A. and T. Zha, 1995. "Error Bands for Impulse Responses," *Federal Reserve Bank of Atlanta Working Paper 95-6*.

__________, 1996. "Does Monetary Policy Generate Recessions?" *manuscript* (Yale University).

Stock, J.H. and M.W. Watson, 1989. "Interpreting the Evidence on Money-Income Causality," *Journal of Econometrics 40*, 161-181.

Thoma, M.A., 1994. "Subsample Instability and Asymmetries in Money-Income Causality," *Journal of Econometrics 64*, 279-306.

Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*, Robert E. Krieger Publishing Company.

Table 1 ML Estimates of Contemporaneous Coefficients

Variables	R	y	P	M
Money Demand	2.89	-82.77	-82.77	82.77
Production Sector	0	377.85	134.15	0
	0	0	725.21	0
Money Supply	0	0	0	242.22

The chi-square likelihood ratio statistic with 10 degrees of freedom is 7.8962.

Figure 1 Impulse Responses

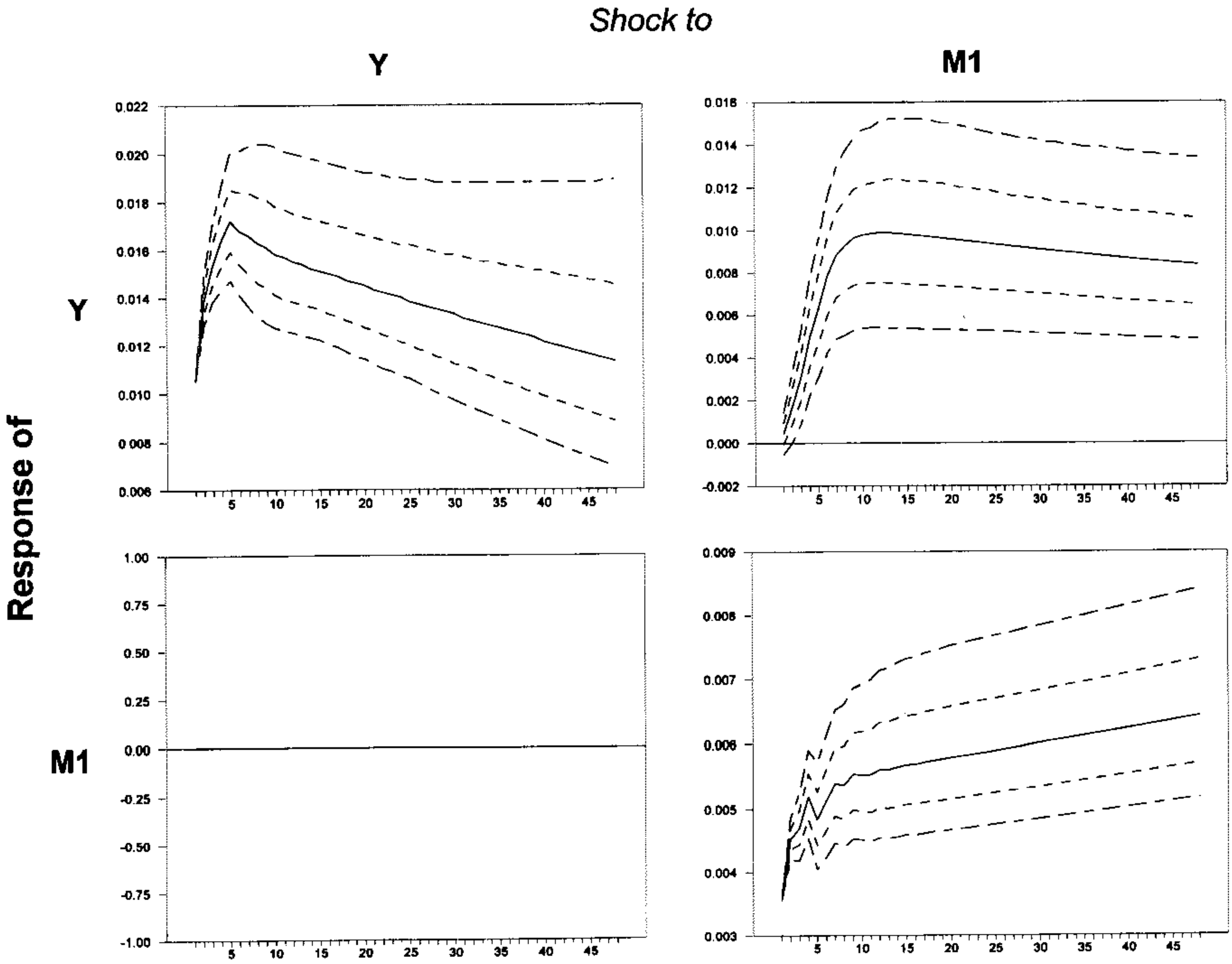

Figure 2 Impulse Responses

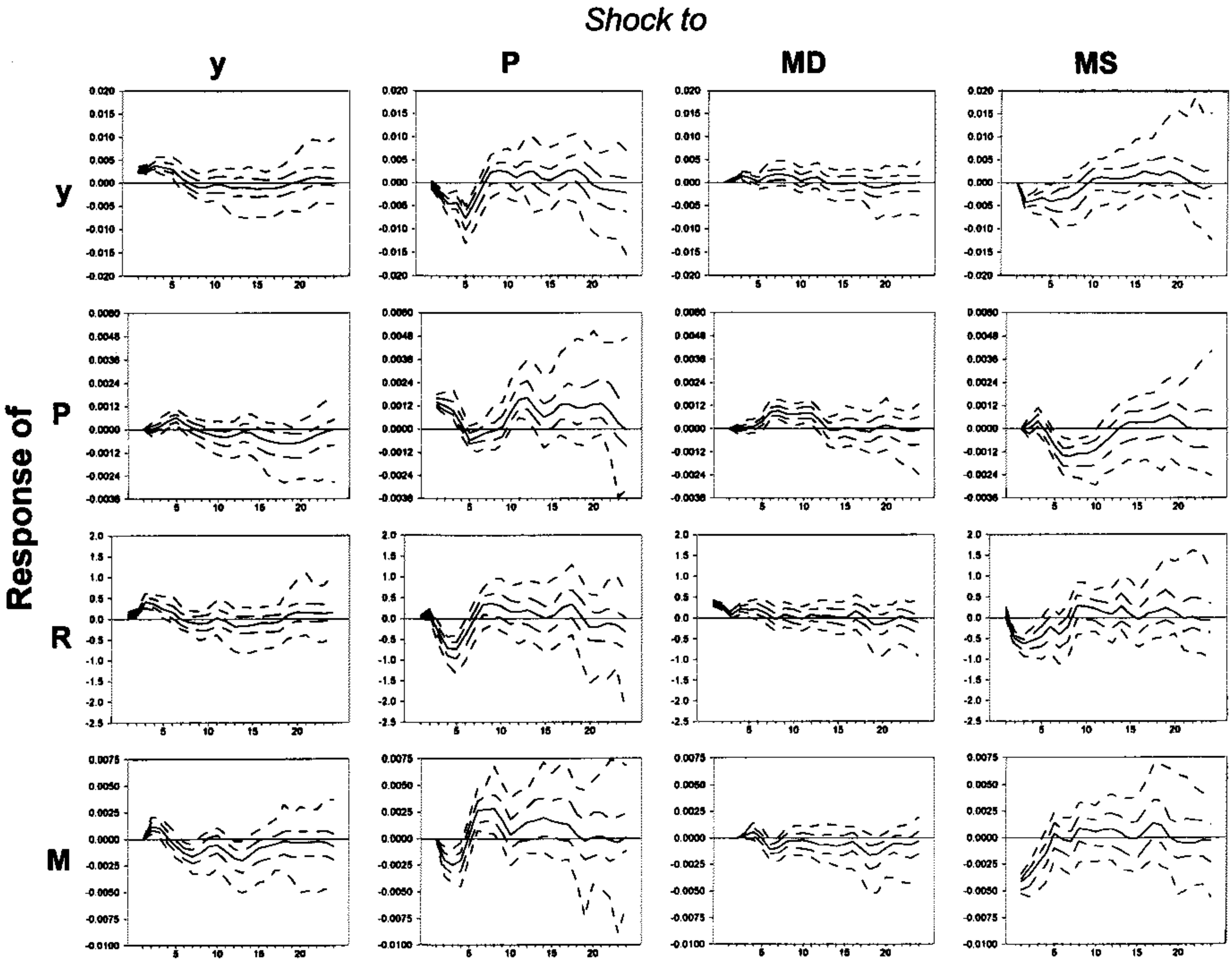

Figure 3 Historical Decompositions of Output