THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies (2010) **77**, 665–696 0034-6527/10/00411011$02.00

doi: 10.1111/j.1467-937X.2009.00578.x

Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference

JUAN F. RUBIO-RAMÍREZ
Duke University and Federal Reserve Bank of Atlanta

DANIEL F. WAGGONER
Federal Reserve Bank of Atlanta

and

TAO ZHA
Federal Reserve Bank of Atlanta and Emory University

First version received October 2005; *final version accepted May* 2009 *(Eds.)*

Structural vector autoregressions (SVARs) are widely used for policy analysis and to provide stylized facts for dynamic stochastic general equilibrium (DSGE) models; yet no workable rank conditions to ascertain whether an SVAR is globally identified have been established. Moreover, when nonlinear identifying restrictions are used, no efficient algorithms exist for small-sample estimation and inference. This paper makes four contributions towards filling these important gaps in the literature. First, we establish general rank conditions for global identification of both identified and exactly identified models. These rank conditions are sufficient for general identification and are necessary and sufficient for exact identification. Second, we show that these conditions can be easily implemented and that they apply to a wide class of identifying restrictions, including linear and certain nonlinear restrictions. Third, we show that the rank condition for exactly identified models amounts to a straightforward counting exercise. Fourth, we develop efficient algorithms for small-sample estimation and inference, especially for SVARs with nonlinear restrictions.

1. INTRODUCTION

Since the seminal work by Sims (1980), identification of structural vector autoregressions (SVARs) has been an unresolved theoretical issue. Filling this theoretical gap is of vital importance because impulse responses based on SVARs have been widely used for policy analysis and to provide stylized facts for dynamic stochastic general equilibrium (DSGE) models (see, for example, Galí, 1999; Smets and Wouters, 2003; Fernández-Villaverde and Rubio-Ramírez, 2004; Christiano *et al.*, 2005; Sims and Zha, 2006*a*).

Drawing inferences from the reduced-form representation that summarizes the sampling information in the data set back to the underlying structural parameters in the SVAR is the root

of the identification problem. In general, making such inferences is impossible, as the reduced-form representation does not give sufficient information to pin down the structural parameters. Therefore, econometricians rely on economic arguments to come up with additional *a priori* restrictions, which are called "identifying restrictions".

Consider a simple one-lag SVAR studied by Leeper *et al.* (1996):

$$\begin{aligned} a_{11}\Delta \log P_{c,t} + a_{31}R_t &= c_1 + b_{11}\Delta \log P_{c,t-1} + b_{21}\Delta \log Y_{t-1} + b_{31}R_{t-1} + \varepsilon_{1,t}, \\ a_{12}\Delta \log P_{c,t} + a_{22}\Delta \log Y_t &= c_2 + b_{12}\Delta \log P_{c,t-1} + b_{22}\Delta \log Y_{t-1} + b_{32}R_{t-1} + \varepsilon_{2,t}, \\ a_{13}\Delta \log P_{c,t} + a_{23}\Delta \log Y_t + a_{33}R_t &= c_3 + b_{13}\Delta \log P_{c,t-1} + b_{23}\Delta \log Y_{t-1} + b_{33}R_{t-1} + \varepsilon_{3,t}, \end{aligned}$$

where a_{ij}, c_j, and b_{ij} for $i, j = 1, 2, 3$ are nonzero coefficients; $\varepsilon_{j,t}$s are uncorrelated random shocks, each of which has an i.i.d. standard normal distribution; $P_{c,t}$ is the price index of commodities, Y_t output, and R_t the nominal short-term interest rate. The first equation is the monetary policy equation in which the interest rate responds sluggishly to output because the data on output are unavailable within the period (Sims and Zha, 2006*a*). Monetary policy can, however, respond to commodity prices, as the data on developments in commodity markets are available daily. In addition, we impose an identifying restriction that monetary policy has no long-run effect on output. This widely used restriction is based on the economic theory that monetary policy has long-run neutrality. The second equation characterizes the behaviour of finished-goods producers. Output responds to commodity prices instantly because commodities are used as inputs in output production, but it responds sluggishly to changes in the interest rate as argued in Christiano *et al.* (2005). Thus a contemporaneous exclusion restriction exists in the second equation. There are no identifying restrictions on the third equation, as this equation reflects the fact that commodity prices are set in active competitive markets and thus respond to all disturbances in the economy (Sims and Zha, 2006*a*). By counting, we have a total of three identifying restrictions.

Now consider an alternative model in which a shock in commodity markets has no long-run effect on output, but monetary policy has a long-run effect on output to allow for the possibility of non-neutrality. By counting, this alternative model has a total of three identifying restrictions, the same number of restrictions as the previous model. The question is: Which model in this example is globally identified?

The standard criterion for identification of SVARs is a necessary condition given by Rothenberg (1971), called the "order condition". The order condition is implemented by simply counting the number of restrictions. For an SVAR to be identified, at least $n(n-1)/2$ restrictions must exist, where n is the number of endogenous variables. In the previous example in which $n = 3$, this condition is satisfied because the number of restrictions is exactly $n(n-1)/2 = 3$. Since Rothenberg's (1971) condition is only necessary, the question is whether the model is globally identified. Many authors have given sufficient conditions for global identification in the framework of traditional simultaneous-equation models.[1] If a researcher mechanically applies these conditions to the identification of SVARs, where the structural covariance matrix is an identity matrix, the identification problem arises again. As argued by Bekker and Pollock (1986) and Leeper *et al.* (1996), linear restrictions on the covariance matrix of shocks generally imply nonlinear restrictions on structural parameters.[2] Therefore, checking

1. Rothenberg (1971) gives general sufficient conditions for global identification for certain types of restrictions on traditional simultaneous-equation models; Dhrymes (1978), Hsiao (1983), and Dhrymes (1994), among others, give other rank conditions for traditional simultaneous-equation models.

2. Exceptions can be found in Hausman and Taylor (1983) and Bekker and Pollock (1986), who show that when a simultaneous-equation model has a recursive (or decomposable) structure, linear restrictions on the covariance matrix lead to linear restrictions on structural parameters as well.

whether an SVAR is globally identified is equivalent to checking whether a system of nonlinear restrictions on the structural parameters has a unique solution, which is in general a difficult problem.

Our approach to global identification of SVARs is different. By exploiting the structure of orthogonal matrices, this alternative approach reduces the problem to counting the rank of certain matrices; in some cases, it further reduces the problem to simply counting the number of restrictions in each equation. Our approach applies to both linear restrictions such as those in the above model and nonlinear restrictions such as those imposed on impulse responses. In particular, we provide rank conditions for global identification of both identified and exactly identified SVARs. We also provide a rank condition for partial identification. These rank conditions are sufficient for general identification and are necessary and sufficient for exact identification.

The existing sufficient rank conditions for identification of SVARs, as discussed in Giannini (1992) and Hamilton (1994, pp. 332–335), apply to local identification only, are mainly designed to analyse linear restrictions on the structural parameters, and can only be numerically verified at a particular point in the parameter space. In contrast, our sufficient condition extends the work of Fisher (1966, Chapters 3 and 4), Hausman and Taylor (1983), and Bekker and Pollock (1986) to global identification. Our theory applies not only to linear restrictions on structural parameters but also to certain nonlinear restrictions on the structural parameters, such as equality restrictions imposed directly on impulse responses. More importantly, we establish that if our rank condition for global identification is satisfied at an arbitrary point in the parameter space, it will be satisfied almost everywhere. This powerful result gives a simple and efficient way to determine whether the model is globally identified in a large parameter space.

Our necessary and sufficient condition for exact identification also works for both linear and certain nonlinear restrictions, and is nothing more than the original Rothenberg (1971) order condition plus our sufficient condition for identification. The necessary condition of Rothenberg (1971) counts the total number of restrictions. Our necessary and sufficient condition not only counts the number of restrictions but also requires that the restrictions follow a certain pattern, equation by equation.

Because our theoretical results are new and different from those in standard textbooks, in Section 5 we demonstrate how to apply our theorems to a number of SVARs widely used in the literature to determine whether these SVARs are globally identified. In particular, we show in Section 5.4 that the introductory SVAR discussed above is globally identified almost everywhere, while the alternative model is not. We demonstrate, through these applications, that our rank conditions can be implemented either as a straightforward exercise of checking ranks of certain matrices or as a simple counting exercise.

Once the global identification issue has been resolved, the next task involves small-sample estimation and inference of the model. For this purpose, both classical and Bayesian methods often require repeated simulations of structural parameters. Such computation is quite expensive, especially when time-varying SVARs are estimated (Uhlig, 1997; Canova and Gambetti, 2004; Cogley and Sargent, 2005; Primiceri, 2005; Sims and Zha, 2006*b*; Gambetti and Canova, 2008). To address this important issue, we use our theoretical results to derive efficient algorithms for exactly identified models and for models identified with sign restrictions. Obtaining accurate small-sample inferences for many relevant problems would be prohibitively expensive and practically infeasible without these new methods. For example, these efficient algorithms make it possible to estimate models with time-varying features and perform the model comparison, as did Rubio-Ramírez *et al*. (2005).

The rest of the paper is organized as follows: Section 2 presents a framework for SVARs. Section 3 develops a general theory of global identification. Section 4 derives necessary and sufficient conditions for exact identification. Section 5 shows how to apply our theory to some widely used SVAR models, whose identifiability has not been formally established. Section 6 uses our theoretical results to derive efficient algorithms for small-sample estimation and inference. Section 7 concludes.

2. THE FRAMEWORK

Our general framework is laid out as follows. In Section 2.1 we discuss a general class of SVARs. In Section 2.2, following Rothenberg (1971), we define global and local identification for this class of models. In Sections 2.3 and 2.4 we introduce and discuss a wide class of identifying restrictions to which our theory applies; these restrictions encompass those in the literature.

2.1. *The structural model*

The class of SVARs we study has the general form

$$\mathbf{y}_t'\mathbf{A}_0 = \sum_{\ell=1}^{p} \mathbf{y}_{t-\ell}'\mathbf{A}_\ell + \mathbf{c} + \boldsymbol{\varepsilon}_t', \qquad \text{for } 1 \le t \le T, \tag{1}$$

where p is the lag length, T the sample size, $\mathbf{y}_t$ an $n \times 1$ vector of endogenous variables, $\boldsymbol{\varepsilon}_t$ an $n \times 1$ vector of exogenous structural shocks,[3] $\mathbf{A}_\ell$ an $n \times n$ matrix of parameters for $0 \le \ell \le p$, and $\mathbf{c}$ is a $1 \times n$ vector of parameters.[4]

The distribution of $\boldsymbol{\varepsilon}_t$, conditional on past information, is Gaussian with mean zero and covariance matrix $\mathbf{I_n}$, the $n \times n$ identity matrix.[5] The initial conditions, $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$, are taken as given. Let

$$\mathbf{A}_+' = \begin{bmatrix} \mathbf{A}_1' & \ldots & \mathbf{A}_p' & \mathbf{c}' \end{bmatrix} \text{ and } \mathbf{x}_t' = \begin{bmatrix} \mathbf{y}_{t-1}' & \ldots & \mathbf{y}_{t-p}' & 1 \end{bmatrix}$$

for $1 \le t \le T$. The dimension of $\mathbf{A}_+$ is $m \times n$, where $m = np + 1$. The model (1) can be written in compact form as

$$\mathbf{y}_t'\mathbf{A}_0 = \mathbf{x}_t'\mathbf{A}_+ + \boldsymbol{\varepsilon}_t'. \tag{2}$$

The parameters of the structural model are $(\mathbf{A}_0, \mathbf{A}_+)$, and we assume that $\mathbf{A}_0$ is invertible. We denote the set of all structural parameters by $\mathbb{P}^S$. The set $\mathbb{P}^S$ is an open dense subset of $\mathbb{R}^{(n+m)n}$. The reduced-form representation implied by the structural model (2) is

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \mathbf{u}_\mathbf{t}',$$

where $\mathbf{B} = \mathbf{A}_+\mathbf{A}_0^{-1}$, $\mathbf{u}_\mathbf{t}' = \boldsymbol{\varepsilon}_t'\mathbf{A}_0^{-1}$, and $E[\mathbf{u_t}\mathbf{u_t}'] = \boldsymbol{\Sigma} = (\mathbf{A}_0\mathbf{A}_0')^{-1}$. The parameters of the reduced-form model are $(\mathbf{B}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a symmetric and positive definite matrix. We

3. In this paper, structural shocks are assumed to be fundamental with respect to the set of variables used for estimation of the SVAR. A different type of identification issue, not addressed in this paper, relates to a choice of variables to be included in the SVAR.

4. This analysis can be easily extended to include exogenous variables other than a constant.

5. Because only the first and second moments are used for identification, the important assumption is that the reduced-form shocks $\mathbf{u}_\mathbf{t}' = \boldsymbol{\varepsilon}_t'\mathbf{A}_0^{-1}$ form a family of distributions uniquely parameterized by their means and variances.

denote the set of all reduced-form parameters by $\mathbb{P}^R$. The set $\mathbb{P}^R$ is an $nm + n(n+1)/2$ dimensional sub-manifold of $\mathbb{R}^{(n+m)n}$ but can be mapped, using the Cholesky decomposition of $\mathbf{\Sigma}$, to an open subset of $\mathbb{R}^{nm+n(n+1)/2}$.

For future reference, define $g : \mathbb{P}^S \to \mathbb{P}^R$ by $g(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{A}_+\mathbf{A}_0^{-1}, (\mathbf{A}_0\mathbf{A}_0')^{-1})$, which defines the relationship between the structural and reduced-form parameters.

2.2. *Identification*

Following Rothenberg (1971), we state that two parameter points, $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$, are observationally equivalent if and only if they imply the same distribution of $\mathbf{y}_t$ for $1 \leq t \leq T$. For the linear Gaussian models of the type studied in this paper, this statement is equivalent to saying that two parameter points are observationally equivalent if and only if they have the same reduced-form representation $(\mathbf{B}, \mathbf{\Sigma})$. The structure of the function g defined above shows that two parameters points $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+)$ have the same reduced form representation if and only if there is an orthogonal matrix $\mathbf{P}$ such that $\mathbf{A}_0 = \tilde{\mathbf{A}}_0\mathbf{P}$ and $\mathbf{A}_+ = \tilde{\mathbf{A}}_+\mathbf{P}$. Global and local identification is characterized in terms of observational equivalence as follows:

Definition 1. *A parameter point* $(\mathbf{A}_0, \mathbf{A}_+)$ *is globally identified if and only if there is no other parameter point that is observationally equivalent.*

Definition 2. *A parameter point* $(\mathbf{A}_0, \mathbf{A}_+)$ *is locally identified if and only if there is an open neighbourhood about* $(\mathbf{A}_0, \mathbf{A}_+)$ *containing no other observationally equivalent parameter point.*

Because observational equivalence is the same as finding an orthogonal matrix $\mathbf{P}$ such that $\mathbf{A}_0 = \tilde{\mathbf{A}}_0\mathbf{P}$ and $\mathbf{A}_+ = \tilde{\mathbf{A}}_+\mathbf{P}$, the set of all $n \times n$ orthogonal matrices plays a central role in our analysis. Following the usual convention, we denote the set of all $n \times n$ orthogonal matrices by $O(n)$.

2.3. *Identifying restrictions*

It is well known that an unrestricted SVAR is neither globally nor locally identified. To achieve identification, therefore, we must restrict the structural parameters. In this section we formally define a broad range of equality restrictions to which our identification theory will apply. These restrictions include widely used linear and nonlinear restrictions studied in the SVAR literature.

Two main classes of equality restrictions are discussed in the literature. The first class identifies the model's structural parameters by imposing linear restrictions on $(\mathbf{A}_0, \mathbf{A}_+)$. This class includes the triangular identification as described by Christiano *et al.* (1996) and the non-triangular identification of Sims (1986), King *et al.* (1991), Gordon and Leeper (1994), Bernanke and Mihov (1998), Zha (1999), and Sims and Zha (2006*b*).

The second class concerns nonlinear restrictions on the structural parameters and includes restrictions imposed on impulse responses, such as short- and long-run restrictions studied by Blanchard and Quah (1993) and Galí (1992). Although these restrictions are nonlinear on the structural parameters, they are linear on the set of impulse responses. Consider long-run impulse responses as an example. When the i-th variable of the structural model is in first difference, the long-run impulse response of the i-th variable to the j-th shock is the element

in row i and column j of [6]

$$\mathbf{IR}_\infty(\mathbf{A}_0, \mathbf{A}_+) = \left(\mathbf{A}_0' - \sum_{\ell=1}^{p} \mathbf{A}_\ell'\right)^{-1}. \tag{3}$$

As one can see, long-run impulse responses are a nonlinear transformation of the underlying structural parameters. Blanchard and Quah (1993) identify their two-variable model, imposing the restriction that a demand shock has no long-run effect on output. This restriction is equivalent to setting an appropriate element of $\mathbf{IR}_\infty$ to zero.

In general, we impose linear restrictions on a transformation of the structural parameter space into the set of $k \times n$ matrices for some $k > 0$. We denote the transformation by $f(\cdot)$ and its domain by $U \subset \mathbb{P}^S$. In the above example, $k = n$ and the domain is the set of all structural parameters such that $\mathbf{A}_0' - \sum_{\ell=1}^{p} \mathbf{A}_\ell'$ is invertible. Linear restrictions on the columns of the transformed space can be represented by $k \times k$ matrices $\mathbf{Q}_j$ for $1 \leq j \leq n$. The parameters $(\mathbf{A}_0, \mathbf{A}_+)$ satisfy the identifying restrictions if and only if

$$\mathbf{Q}_j f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e_j} = 0, \text{ for } 1 \leq j \leq n, \tag{4}$$

where $\mathbf{e_j}$ is the j-th column of the $n \times n$ identity matrix $\mathbf{I}_n$. We denote the number of restrictions on the j-th column by q_j. Thus, the rank of $\mathbf{Q}_j$ is q_j, and the total number of restrictions is $\sum_{j=1}^{n} q_j = q$. Because the ordering of the columns of $f(\cdot)$ is completely arbitrary, we assume without loss of generality that

$$q_1 \geq q_2 \geq \ldots \geq q_n. \tag{5}$$

This assumption allows a simpler statement of our results.

The transformation $f(\cdot)$ must satisfy certain conditions. The most important one is that right multiplication by an orthogonal matrix must commute with the transformation. In the above example, if $\mathbf{P}$ is orthogonal, then

$$\begin{aligned}\mathbf{IR}_\infty(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) &= \left((\mathbf{A}_0\mathbf{P})' - \sum_{\ell=1}^{p} (\mathbf{A}_\ell\mathbf{P})'\right)^{-1} \\ &= \left(\mathbf{A}_0' - \sum_{\ell=1}^{p} \mathbf{A}_\ell'\right)^{-1} (\mathbf{P}')^{-1} = \mathbf{IR}_\infty(\mathbf{A}_0, \mathbf{A}_+)\mathbf{P}.\end{aligned}$$

We call transformations that have this property "admissible" ones, as formally defined below.

Condition 1. *The transformation $f(\cdot)$, with the domain U, is admissible if and only if for any $\mathbf{P} \in O(n)$ and $(\mathbf{A}_0, \mathbf{A}_+) \in U$, $f(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) = f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{P}$.*

Other regularity conditions on $f(\cdot)$, required by various proofs of our theoretical results (see the appendices), are given below.

Condition 2. *The transformation $f(\cdot)$, with the domain U, is regular if and only if U is open and f is continuously differentiable with $f'(\mathbf{A}_0, \mathbf{A}_+)$ of rank kn for all $(\mathbf{A}_0, \mathbf{A}_+) \in U$.*

6. As Blanchard and Quah (1993) point out, if the i-th variable is in first differences and $\mathbf{A}(L)^{-1} = \sum_{k=0}^{\infty} \mathbf{C_k} L^k$; where $\mathbf{A}(L) = \mathbf{A}_0' - \sum_{\ell=1}^{p} \mathbf{A}_\ell' L^\ell$, then the long run response of the i-th variable to the j-th shock is the element in row i and column j of the matrix $\sum_{k=0}^{\infty} \mathbf{C}_k$. Equation (3) follows by substituting $L = 1$ in $\mathbf{A}(L)$.

Condition 3. *The transformation $f(\cdot)$, with the domain U, is strongly regular if and only if it is regular and $f(U)$ is dense in the set of $k \times n$ matrices.*

To see how to check whether Conditions 2.3 and 2.3 are satisfied, consider the above example represented by (3). In this example, $k = n$. The domain of $\mathbf{IR}_\infty$ is the set of all structural parameters such that $\mathbf{A}_0' - \sum_{\ell=1}^{p} \mathbf{A}_\ell'$ is invertible; therefore, this domain is an open set. Because $\mathbf{IR}_\infty$ is the composition of the linear projection that projects $(\mathbf{A}_0, \mathbf{A}_+)$ to $\mathbf{A}_0' - \sum_{\ell=1}^{p} \mathbf{A}_\ell'$ and the invertible function that maps a matrix to its inverse, $\mathbf{IR}_\infty$ is continuously differentiable, and its derivative is of rank $nk = n^2$. According to Condition 2.3, the transformation $\mathbf{IR}_\infty$ is regular. Moreover, the image of $\mathbf{IR}_\infty$ is equal to the set of all invertible matrices and thus is dense. According to Condition 2.3, $\mathbf{IR}_\infty$ is strongly regular.

For admissible transformations, it is important to note that the restrictions given by (4) alone are insufficient for global identification. To see this point, suppose that $\mathbf{D}$ is any $n \times n$ diagonal matrix with plus or minus ones along the diagonal. Such matrices are orthogonal and thus commute with the transformation $f(\cdot)$. Since $\mathbf{De_j} = \pm\mathbf{e_j}$, if $(\mathbf{A}_0, \mathbf{A}_+)$ satisfy (4), then $(\mathbf{A}_0\mathbf{D}, \mathbf{A}_+\mathbf{D})$ will also satisfy (4). The system is thus not identified. Consequently, one must choose a normalization rule to determine the sign of each equation, as in standard textbooks (e.g. Dhrymes, 1978, p. 284; Greene, 1993, p. 590). While a poor choice can distort statistical inference of impulse responses (Waggoner and Zha, 2003; Hamilton *et al.*, 2007), the theory developed in this paper works for any choice of normalization. A general definition of normalization is given below.

Definition 3. *A normalization rule can be characterized by a set $N \subset \mathbb{P}^S$ such that for any structural parameter point $(\mathbf{A}_0, \mathbf{A}_+) \in \mathbb{P}^S$, there exists a unique $n \times n$ diagonal matrix $\mathbf{D}$ with plus or minus ones along the diagonal such that $(\mathbf{A}_0\mathbf{D}, \mathbf{A}_+\mathbf{D}) \in N$.*

The set N is the collection of normalized structural parameters. Throughout this paper we assume that all SVAR models are normalized via some normalization rule N. Using the transformation $f(\cdot)$, the sets U and N, and the matrices $\mathbf{Q}_j$ for $1 \leq j \leq n$, we are in a position to specify the set of restricted parameters as

$$R = \{(\mathbf{A}_0, \mathbf{A}_+) \in U \cap N \mid \mathbf{Q}_j f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e_j} = 0 \text{ for } 1 \leq j \leq n\}. \tag{6}$$

If the transformation $f(\cdot)$ is admissible, regular, or strongly regular, then we say that the restrictions given by (6) are admissible, regular, or strongly regular, respectively.

2.4. *Examples of identifying restrictions*

In the previous subsection, we have studied an example of restrictions on long-run impulse responses and shown that they are admissible and strongly regular. In this subsection, we show that the restrictions are admissible and strongly regular in other well-established examples.

Example 1. *Linear restrictions on $\mathbf{A}_0$ or $(\mathbf{A}_0, \mathbf{A}_+)$. For linear restrictions on the contemporaneous coefficient matrix $\mathbf{A}_0$, $f(\cdot)$ is a projection onto $\mathbf{A}_0$ and hence $f(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{A}_0$. More generally, for linear restrictions on both contemporaneous and lagged (predetermined) coefficients, the transformation $f(\cdot)$ is simply the following identity mapping:*

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_+ \end{bmatrix}.$$

Clearly, such restrictions are admissible and strongly regular.

Example 2. *Short-run restrictions on impulse responses. The impulse response of the i-th variable to the j-th shock at horizon h corresponds to the element in row i and column j of the following matrix:*

$$\mathbf{IR}_h = \left(\mathbf{A}_0^{-1}\mathbf{J}'\mathbf{F}^{\mathbf{h}}\mathbf{J}\right)',$$

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}_1\mathbf{A}_0^{-1} & \mathbf{I}_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{p-1}\mathbf{A}_0^{-1} & 0 & \cdots & \mathbf{I}_n \\ \mathbf{A}_p\mathbf{A}_0^{-1} & 0 & \cdots & 0 \end{bmatrix} \quad \textit{and} \quad \mathbf{J} = \begin{bmatrix} \mathbf{I}_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Let $f(\cdot)$ be the transformation that maps $(\mathbf{A}_0, \mathbf{A}_+)$ to $[\mathbf{IR}_0' \ \ldots \ \mathbf{IR}_p']'$. Clearly $f(\cdot)$ is continuously differentiable, its domain is $\mathbb{P}^S$, and its image consists of all $(n+m) \times n$ matrices for which the first $n \times n$ block is invertible. Clearly, the domain is open and the image is dense. Showing that the function $f(\cdot)$ has a continuously differentiable inverse is not difficult, but it is tedious. This result implies that the rank of the derivative of $f(\cdot)$ is $(n+m)n$. Thus, linear restrictions on impulse responses are admissible and strongly regular.

Example 3. *If one wishes to have restrictions on the structural parameters and on both short- and long-run impulse responses, one can combine the transformations described above. Such a combination will always be admissible, and it will be strongly regular as long as the rank of the derivative is kn. In Galí (1992), for example, identifying restrictions involve both contemporaneous and long-run impulse responses. Here $k = 2n$ and*

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{IR}_0 \\ \mathbf{IR}_\infty \end{bmatrix}.$$

In the example discussed in this paper's introduction, $n = 3$, $k = 2n = 6$, and

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{IR}_\infty \end{bmatrix}.$$

Verifying that both transformations are admissible and strongly regular is a straightforward exercise.

3. GLOBAL IDENTIFICATION

In this section we present our theory for global identification of SVARs with both linear and certain nonlinear restrictions. In Section 3.1 we derive rank conditions that are sufficient for global identification. In Section 3.2 we extend our results to partial identification. Finally, in Section 3.3 we analyse identification almost everywhere.

3.1. *A rank condition for global identification*

We now develop a sufficient condition for global identification. This rank condition is very general, and in Section 5 we show how to apply it to a number of widely used models in the literature.

The following matrix is the key to the establishment of our rank condition. For $1 \leq j \leq n$ and any $k \times n$ matrix $\mathbf{X}$, we define $\mathbf{M}_j(\mathbf{X})$ by

$$\underset{(k+j)\times n}{\mathbf{M}_j(\mathbf{X})} = \begin{bmatrix} \underset{k\times k}{\mathbf{Q}_j} \; \underset{k\times n}{\mathbf{X}} \\ \begin{bmatrix} \underset{j\times j}{\mathbf{I}} & \underset{j\times(n-j)}{\mathbf{0}} \end{bmatrix} \end{bmatrix}.$$

We state and prove the following key theorem:

Theorem 1. *Consider an SVAR with admissible restrictions represented by R. If $(\mathbf{A}_0, \mathbf{A}_+) \in R$ and $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ is of rank n for $1 \leq j \leq n$, then the SVAR is globally identified at the parameter point $(\mathbf{A}_0, \mathbf{A}_+)$.*

Proof. To prove the theorem, it suffices to show that if the SVAR is not identified at $(\mathbf{A}_0, \mathbf{A}_+)$, then there exists a j such that $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ is of rank strictly less than n. From the discussion in Section 2.2 and the fact that a normalization is used, one can conclude that if the SVAR is not identified at $(\mathbf{A}_0, \mathbf{A}_+)$, then there exists a $\mathbf{P} = (p_{i,j}) \in O(n)$ such that $\mathbf{P}$ is not diagonal and $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R$. Let j be the index of the first column of $\mathbf{P}$ that has a non-zero off-diagonal element. We shall show that the rank of $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ is strictly less than n.

Let $\mathbf{v_j} = \mathbf{P}\mathbf{e_j} - p_{j,j}\mathbf{e_j}$, where $\mathbf{e_j}$ is the j-th column of $\mathbf{I}_n$. Since $\mathbf{v_j} \neq 0$, to complete the proof it suffices to show $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))\mathbf{v_j} = 0$. Because both $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ are in R, $\mathbf{Q}_j f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{v_j} = 0$. Thus, the upper block of $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))\mathbf{v_j}$ is zero. Because $\mathbf{P}$ is orthogonal and j is the index of the first column of $\mathbf{P}$ that has a non-zero off-diagonal element, the first $j-1$ elements of $\mathbf{v_j}$ are zero, and the j-th element of $\mathbf{v_j}$ is zero by construction. This result implies that the lower block of $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))\mathbf{v_j}$ is also zero. Thus $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))\mathbf{v_j} = 0$ as required. ||

Since the ordering of columns of $f(\cdot)$ is arbitrary, and the condition in Theorem 1 may be satisfied under one ordering but not under another, one might wish to experiment with all possible orderings when using the rank condition. Choosing correct ordering, however, can eliminate an unnecessary search and make it more efficient to check the rank condition. It can be shown that, except for a set of measure zero in $\mathbb{P}^S$, our ordering convention given by (5) will produce $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ of rank n for $1 \leq j \leq n$ if there is any ordering of the columns such that $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ is of rank n for $1 \leq j \leq n$.

3.2. *Partial identification*

Instead of identifying the entire system, researchers are sometimes interested in identifying only a single equation or a subset of equations in the system.[7] The theoretical results in the previous subsection can be extended to this partial-identification case. We begin with a formal definition of partial identification for SVAR models.

Definition 4. *The j-th equation is globally identified at the parameter point $(\mathbf{A}_0, \mathbf{A}_+) \in R$ if and only if there does not exist another observationally equivalent parameter point $(\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_+) \in R$ such that $\mathbf{A}_0\mathbf{e_j} \neq \tilde{\mathbf{A}}_0\mathbf{e_j}$ or $\mathbf{A}_+\mathbf{e_j} \neq \tilde{\mathbf{A}}_+\mathbf{e_j}$, where $\mathbf{e_j}$ is the j-th column of the $n \times n$ identity matrix.*

7. We thank an anonymous referee for suggesting this important generalization.

Global identification says that one cannot have distinct observationally equivalent parameter points that satisfy the restrictions. Partial identification allows for observationally equivalent parameter points that satisfy the restrictions, but these observationally equivalent parameter points cannot differ in the j-th equation. A subset of equations is identified if each equation in the subset is identified. The following theorem gives a sufficient rank condition for partial identification.

Theorem 2. *Consider an SVAR with admissible restrictions represented by R. For $(\mathbf{A}_0, \mathbf{A}_+) \in R$, if $\mathbf{M}_i(f(\mathbf{A}_0, \mathbf{A}_+))$ is of rank n for $1 \leq i \leq j$, then the j-th equation is globally identified at the parameter point $(\mathbf{A}_0, \mathbf{A}_+)$.*

Proof. The proof follows directly from the proof of Theorem 1. ||

As in Theorem 1, the condition in Theorem 2 may be satisfied under one ordering of the equations but not under another. As mentioned above, checking the condition under our ordering convention is sufficient for almost all structural parameters.

3.3. *Global identification almost everywhere*

In the SVAR literature, the existing rank conditions for local identification are (numerically) checked at a particular parameter point. The chosen point is typically the estimate of the model parameters. Knowing whether the model is identified at different points in the parameter space prior to the estimation step is often important. Here, we develop two results to answer this question. We begin with the definition of the set

$$K = \left\{(\mathbf{A}_0, \mathbf{A}_+) \in R \mid \text{rank}\left(\mathbf{M}_j\left(f\left(\mathbf{A}_0, \mathbf{A}_+\right)\right)\right) = n \text{ for } 1 \leq j \leq n\right\}. \tag{7}$$

According to Theorem 1, if the restrictions are admissible, the model will be globally identified on the set K. It can be easily shown that the set K is open. Thus, if the structural parameter point $(\mathbf{A}_0, \mathbf{A}_+) \in R$ satisfies the rank condition, then there exists a neighbourhood around $(\mathbf{A}_0, \mathbf{A}_+)$ such that all the structural parameters within that neighbourhood satisfy the rank condition. The implication of this result is that if the model is identified at the estimated value of the structural parameters, there is no need to check if it is identified at nearby points. The next theorem gives an even stronger result: if the restrictions are admissible and regular and if the model is globally identified at any one point in the structural parameter space, then the model is, in fact, globally identified almost everywhere.

Theorem 3. *Consider an SVAR with admissible and regular restrictions represented by R. Either K is empty or the complement of K in R is of measure zero in R.*

Proof. The proof is provided in Appendix A. ||

This theorem demonstrates that one can randomly choose an element of R and then verify the rank condition. If the rank condition is satisfied, we know that the model is globally identified almost everywhere.

4. EXACT IDENTIFICATION

Since much of the SVAR literature involves exactly identified models, in this section we show how our sufficient condition for global identification becomes a necessary and sufficient condition for exact identification.

4.1. *Definition*

We begin with a definition of exact identification.

Definition 5. *Consider an SVAR with restrictions represented by R. The SVAR is exactly identified if and only if, for almost any reduced-form parameter point* $(\mathbf{B}, \mathbf{\Sigma})$*, there exists a unique structural parameter point* $(\mathbf{A}_0, \mathbf{A}_+) \in R$ *such that* $g(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{B}, \mathbf{\Sigma})$.

This definition differs slightly from that of Hamilton (1994, p. 250). The subtle difference is crucial to understanding SVARs with non-triangular restrictions on $\mathbf{A}_0$. Hamilton (1994, p. 250) defines exact identification by insisting that for any (not just for almost any) point in the reduced-form parameter space, there exists a unique set of structural parameters that imply these reduced-form parameters. The following theorem reveals that Hamilton's definition precludes all SVAR models with non-triangular restrictions on $\mathbf{A}_0$ from being exactly identified. The restrictions are triangular if there is a linear transformation of the variables and a permutation of the equations such that the transformed contemporaneous matrix is always triangular. Triangular restrictions, in a slightly more general context, will again be considered in Section 6.3.

Theorem 4. *Consider an SVAR with restrictions represented by R. Let the transformation* $f(\cdot)$ *be given by* $f(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{A}_0$ *so that the SVAR is identified via linear restrictions on the contemporaneous matrix* $\mathbf{A}_0$*. If for every reduced-form parameter point* $(\mathbf{B}, \mathbf{\Sigma})$ *there exists a unique structural parameter point* $(\mathbf{A}_0, \mathbf{A}_+) \in R$ *such that* $g(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{B}, \mathbf{\Sigma})$*, then the restrictions on* $\mathbf{A}_0$ *are triangular.*

Proof. The proof is provided in Appendix C. ||

Thus, Definition 4.1, not Hamilton's (1994) original definition, is needed to allow for the possibility that non-triangular models may be exactly identified.

4.2. *Existence of a unique rotation matrix*

Definition 4.1 deals with the reduced form parameters. The following theorem gives an equivalent formulation of exact identification in terms of structural parameters.

Theorem 5. *Consider an SVAR with restrictions represented by R. The SVAR is exactly identified if and only if, for almost every structural parameter point* $(\mathbf{A}_0, \mathbf{A}_+) \in U$*, there exists a unique matrix* $\mathbf{P} \in O(n)$ *such that* $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R$.

Proof. The proof is provided in Appendix A. ||

Theorem 5 forms the basis of efficient classical and Bayesian small-sample methods for estimation and inference. For an exactly identified SVAR, Theorem 5 states that there exists a unique orthogonal matrix $\mathbf{P}$ such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ satisfies the restrictions for almost every unrestricted structural parameter point $(\mathbf{A}_0, \mathbf{A}_+)$. If one is able to draw unrestricted structural parameters or reduced-form parameters, this result gives us a practical way to find the set of structural parameters that satisfy the identifying restrictions. For each draw of unrestricted structural parameters, one needs to find only an orthogonal matrix $\mathbf{P}$ that rotates the unrestricted draw to the one that satisfies the restrictions.[8] If the original draw is from the reduced-form

8. This procedure applies to the maximum likelihood estimation as well. One first obtains a maximum likelihood estimate of the parameters in an unidentified system and then uses $\mathbf{P}$ to rotate these parameters to get the estimate of the structural parameters that satisfy the identifying restrictions.

parameters ($\mathbf{B}$, $\mathbf{\Sigma}$), one can rotate the Cholesky decomposition of $\mathbf{\Sigma}$ to get a draw that satisfies the restrictions. Building an efficient algorithm to find the needed orthogonal matrix $\mathbf{P}$ is a difficult part of this procedure. In Section 6, we show that such an algorithm exists and, for systems in which the restrictions on $f(\cdot)$ are triangular, we develop an even faster algorithm.

Why is this result important? Consider an exactly identified model with restrictions on impulse responses. The existing methods in the literature typically solve a system of nonlinear equations to find $\mathbf{P}$. This traditional approach becomes very inefficient if a large number of simulations is required to obtain accurate results for small-sample estimation and inference. When time-varying SVARs are studied (Canova and Gambetti, 2004; Cogley and Sargent, 2005; Primiceri, 2005; Sims and Zha, 2006*b*), the traditional approach is practically infeasible because the system of nonlinear equations would have to be solved for each possible state of the structural parameters. In contrast, Theorem 7 implies that the restrictions for exactly identified systems have a special structure that can be exploited to obtain an efficient method for finding the orthogonal matrix $\mathbf{P}$.

4.3. *Rank conditions for exact identification*

We are now able to develop rank conditions for exact identification. The well-known order condition for exact identification implied by Rothenberg (1971) states that the total number of restrictions must be equal to $n(n-1)/2$. In the next theorem, we prove that if Rothenberg's (1971) order condition is satisfied, the sufficient condition in Theorem 1 becomes necessary as well.

Theorem 6. *Consider an SVAR with admissible and strongly regular restrictions represented by R. The SVAR is exactly identified if and only if the total number of restrictions is equal to $n(n-1)/2$ and the rank condition in Theorem 1 is satisfied for some $(\mathbf{A}_0, \mathbf{A}_+) \in R$.*

Proof. The proof is provided in Appendix B. ||

In light of Theorem 3, Theorem 6 gives us a checkable necessary and sufficient condition for exact identification. In what follows, we develop a more powerful condition that does not require calculating the rank of any matrix but simply involves verifying whether the numbers of restrictions satisfy a certain order.

Theorem 7. *Consider an SVAR with admissible and strongly regular restrictions represented by R. The SVAR is exactly identified if and only if $q_i = n - j$ for $1 \leq j \leq n$.*

Proof. The proof is provided in Appendix B. ||

The rank condition in Theorem 7 is a simple counting exercise. The necessary condition of Rothenberg (1971) counts the total number of restrictions. Our necessary and sufficient condition not only counts the number of restrictions but also requires that the restrictions follow a certain equation-by-equation pattern.

4.4. *Local vs. global identification*

In this subsection we analyse a model that is locally identified but not globally identified. The analysis is particularly insightful as it shows that the rank condition in the literature for local identification does not provide any guidance as to whether or not the model is identified globally.

Consider the three-variable SVAR studied by Hausman and Taylor (1983) and Sims and Zha (1999):[9]

$$a_{11}y_{1t} + a_{21}y_{2t} = \varepsilon_{1t},$$
$$a_{22}y_{2t} + a_{32}y_{3t} = \varepsilon_{2t},$$
$$a_{13}y_{1t} + a_{33}y_{3t} = \varepsilon_{3t},$$

where y_{jt} for $j = 1, 2, 3$ is an endogenous variable, and the random shocks ε_{jt}s are serially uncorrelated and have an i.i.d. standard normal distribution. The contemporaneous coefficient matrix for this model can be written as

$$\mathbf{A}_0 = \begin{bmatrix} a_{11} & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ 0 & a_{32} & a_{33} \end{bmatrix}. \tag{8}$$

Since there is only one restriction in each equation, $q_1 = q_2 = q_3 = 1$. The model satisfies Rothenberg's (1971) order condition that the total number of restrictions equals $n(n-1)/2 = 3$. Furthermore, using the procedure outlined in Hamilton (1994, pp. 332–335), one can show that the model is locally identified at the parameter point $a_{11} = a_{22} = a_{33} = 1$ and $a_{13} = a_{21} = a_{32} = 2$.[10] According to Theorem 7, however, this model is not exactly identified at this parameter point.

A natural question is whether this parameter point is so special that the model may still be globally identified at other parameter points. A thorough analysis reveals that the space of reduced-form parameters divides into three sets: a set of positive measure on which the model is locally identified but not globally identified, a set of positive measure on which no representation of structural parameters satisfies the restrictions, and a set of measure zero on which the model is globally identified (see Rubio-Ramírez *et al.* (2008) for a formal proof of this result). In summary, the model is locally identified at every structural parameter point in the parameter space R, but it is not globally identified almost everywhere in R.

To see that the model is not identified at the parameter point $a_{11} = a_{22} = a_{33} = 1$ and $a_{13} = a_{21} = a_{32} = 2$, consider the orthogonal matrix $\mathbf{P}$

$$\mathbf{P} = \begin{bmatrix} 2/3 & 2/3 & -1/3 \\ -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \end{bmatrix}.$$

One can easily show that

$$\tilde{\mathbf{A}}_0 = \mathbf{A}_0\mathbf{P} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}.$$

Thus, $\tilde{\mathbf{A}}_0$ satisfies the restrictions and is observationally equivalent to $\mathbf{A}_0$. This powerful example shows how a structural model can be locally identified but nonetheless fails to be globally identified; it highlights practical distinctions between local identification and global identification.

9. Fubac *et al.* (2007) show that this SVAR can be derived directly from the three-variable forward-looking New-Keynesian model studied by Cochrane (2006).

10. The rank condition for local identification in Hamilton (1994, pp. 332–335) can be numerically checked at any parameter point that satisfies the identifying restrictions. Alternatively, one can use Rothenberg's (1971) information matrix to check if the model is locally identified at a particular parameter point, as suggested by Sargent (1976) and recently employed by Iskrev (2007).

5. APPLICATION

Since our theoretical results, particularly those concerning nonlinear restrictions, are new, we demonstrate in this section how to apply our theory to a number of existing SVARs studied in the literature. For most of these models, global identification has not been formally established.

5.1. *A triangular SVAR*

If the restrictions on $\mathbf{A}_0$ are triangular and no other restrictions exist, as in Eichenbaum and Evans (1995) and Christiano *et al.* (1996), the model is exactly identified. Applying Theorem 7 becomes trivial.

5.2. *A monetary SVAR*

To identify the systematic monetary policy behaviour, Sims (1986), Gordon and Leeper (1994), Bernanke and Mihov (1998), Leeper and Zha (2003), and Sims and Zha (2006*b*), among others, propose identifying restrictions on $(\mathbf{A}_0, \mathbf{A}_+)$. Their approaches focus on an economic interpretation of the parameters in structural equations themselves. In particular, they separate the monetary policy equation from the money demand equation and other non-policy equations. The restrictions require non-triangular relationships between financial variables such as the interest rate and money.

The following $\mathbf{A}_0$ gives a particular example of restrictions on the contemporaneous coefficients only:[11]

$$\mathbf{A}_0 = \begin{array}{c} \\ \log Y \\ \log P \\ R \\ \log M \\ \log P_{\mathrm{c}} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathrm{PS} & \mathrm{PS} & \mathrm{MP} & \mathrm{MD} & \mathrm{Inf} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & 0 & a_{14} & a_{15} \\ 0 & a_{22} & 0 & a_{24} & a_{25} \\ 0 & 0 & a_{33} & a_{34} & a_{35} \\ 0 & 0 & a_{43} & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & a_{55} \end{bmatrix} \end{array}, \tag{9}$$

where the transformation function for this case is

$$f(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{A}_0.$$

The five variables in the model are log gross domestic product (GDP) (log Y), log GDP deflator (log P), the nominal short-term interest rate (R), log M3 (log M), and log commodity prices (log P_{c}). The monetary policy (MP) column in (9) represents a central bank's contemporaneous behaviour; the information (Inf) column describes the commodity (information) market; the MD column corresponds to the money demand equation; and the block consisting of the first two columns represents the production sector (PS), whose variables are arbitrarily ordered to be upper triangular. For this model, we have $k = n = 5$. To use Theorem 1, we express the restriction matrices $\mathbf{Q}_j$ for $j = 1, \ldots, 5$ as

$$\mathbf{Q}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{Q}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{Q}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

11. For an example of identifying restrictions on the lagged structure, see Zha (1999).

$$\mathbf{Q}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It follows from these $\mathbf{Q}_j$s that $q_1 = 4$, $q_2 = 3$, $q_3 = 3$, $q_4 = 1$, and $q_5 = 0$; and the total number of restrictions ($\sum_{j=1}^{5} q_j$) is 11, greater than $n(n-1)/2 = 10$. Therefore, according to Rothenberg's (1971) order condition, the model may be identified. Since Rothenberg's condition is only necessary, we apply the sufficient condition of Theorem 1 by filling the rank matrices $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ for $j = 1, \ldots, 5$ as

$$\mathbf{M}_1 = \begin{bmatrix} 0 & a_{22} & 0 & a_{24} & a_{25} \\ 0 & 0 & a_{33} & a_{34} & a_{35} \\ 0 & 0 & a_{43} & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & a_{55} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} 0 & 0 & a_{33} & a_{34} & a_{35} \\ 0 & 0 & a_{43} & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & a_{55} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_3 = \begin{bmatrix} a_{11} & a_{12} & 0 & a_{14} & a_{15} \\ 0 & a_{22} & 0 & a_{24} & a_{25} \\ 0 & 0 & 0 & 0 & a_{55} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & a_{55} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{M}_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Clearly, there exist values of the a_{ij}s such that the rank of the matrix $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ is $n = 5$ for $j = 1, \ldots, 5$ (for example, take $a_{11} = a_{22} = a_{33} = a_{44} = a_{55} = a_{14} = 1$ and all other parameters to be zero). According to Theorems 1 and 3, the model is globally identified for almost all structural parameters.

5.3. *Restrictions on impulse responses*

Imposing restrictions directly on impulse responses has become increasingly popular (Sims, 2005). To illustrate how to apply our theory to this type of nonlinear restrictions, we follow Peersman and Smets (2003) and consider a four-variable SVAR with three contemporaneous and three long-run restrictions on impulse responses. The four endogenous variables are quarterly output growth ($\Delta \log Y$), quarterly inflation (ΔP), the nominal short-term interest rate (R), and a quarterly change of the nominal exchange rate euro/dollar ($\Delta \log Ex$). The short-run restrictions on impulse responses are:

- Monetary policy shocks have no contemporaneous effect on output.
- Exchange rate shocks have no contemporaneous effect on output.
- Exchange rate shocks have no contemporaneous effect on the interest rate.

TABLE 1
Restrictions implying that the model is identified

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{IR}_0 \\ \mathbf{IR}_\infty \end{bmatrix} =$$

	Ex	MP	D	S
$\Delta \log Y$	0	0	×	×
$\Delta \log P$	×	×	×	×
R	0	×	×	×
$\Delta \log Ex$	×	×	×	×
$\Delta \log Y$	0	0	0	×
$\Delta \log P$	×	×	×	×
R	×	×	×	×
$\Delta \log Ex$	×	×	×	×

TABLE 2
Restrictions implying that the model is only partially identified

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{IR}_0 \\ \mathbf{IR}_\infty \end{bmatrix} =$$

	Ex	MP	D	S
$\Delta \log Y$	0	0	×	×
$\Delta \log P$	×	×	×	0
R	0	×	×	×
$\Delta \log Ex$	×	×	×	×
$\Delta \log Y$	0	×	0	×
$\Delta \log P$	×	×	×	×
R	×	×	×	×
$\Delta \log Ex$	×	×	×	×

The long-run restrictions on impulse responses are:

- Aggregate demand shocks have no long-run effect on output.
- Monetary policy shocks have no long-run effect on output.
- Exchange rate shocks have no long-run effect on output.

The transformation function and the identifying restrictions on the transformed parameters are represented in Table 1. Along the top row of the table, "Ex" stands for a shock to the exchange rate market; "MP" stands for a monetary policy shock; "D" stands for a demand shock; and "S" stands for a supply shock. The symbol "×" means that no restriction is imposed, and "0" means an exclusion restriction.

From this table one can see that $n = 4$, $k = 2n = 8$, $q_1 = 3$, $q_2 = 2$, $q_3 = 1$, and $q_4 = 0$. The total number of restrictions ($\sum_{j=1}^{4} q_j$) is equal to $n(n-1)/2 = 6$ and Rothenberg's (1971) order condition for exact identification holds. Because $q_j = n - j$ for $j = 1, \ldots, 4$, this model is exactly identified according to Theorem 7.

Now consider an alternative identification in which supply shocks have no contemporaneous effect on inflation because of the price stickiness, and monetary policy shocks may have a long run effect on output because of the non-neutrality in money. This alternative identification implies the set of restrictions on $f(\mathbf{A}_0, \mathbf{A}_+)$ represented in Table 2.

For this alternative set of restrictions, the total number of restrictions is still equal to 6, and therefore Rothenberg's (1971) order condition for exact identification holds. But it is easy to see from Theorem 7 that the model is now not exactly identified because $q_1 = 3$ and $q_2 = q_3 = q_4 = 1$.

Is the model partially identified? To address this question, we express the restriction matrices $\mathbf{Q}_j$ and the rank matrices $\mathbf{M}_i(f(\mathbf{A}_0, \mathbf{A}_+))$ for $j = 1, \ldots, 4$ as

$$\mathbf{Q}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{Q}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_4 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_1 = \begin{bmatrix} 0 & 0 & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & 0 & \times \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} 0 & 0 & \times & \times \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_3 = \begin{bmatrix} 0 & \times & 0 & \times \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{M}_4 = \begin{bmatrix} \times & \times & \times & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

These matrices show that although the rank of $\mathbf{M}_j(f(\mathbf{A}_0, \mathbf{A}_+))$ is 4 for $j = 1, 3, 4$ at almost any parameter point, the rank of $\mathbf{M}_2(f(\mathbf{A}_0, \mathbf{A}_+))$ is always 3. Therefore, one can use an analysis similar to that in Section 4.4 to show that the second, third, and fourth equations are not

identified. The system, however, is partially identified because applying Theorem 2 to this case leads to the conclusion that the first equation is identified.

5.4. *A combination of linear and nonlinear restrictions*

We now return to the simple SVAR discussed in the introduction to check whether it is globally identified. The model involves two linear restrictions on the contemporaneous coefficient matrix $\mathbf{A}_0$ and one nonlinear restriction that imposes on the long-run impulse response of output to a monetary policy shock. These restrictions can be expressed as

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{IR}_\infty \end{bmatrix} = \begin{array}{c} \\ \Delta \log P_c \\ \Delta \log Y \\ R \\ \Delta \log P_c \\ \Delta \log Y \\ R \end{array} \begin{array}{c} \begin{array}{ccc} \text{MP} & \text{P} & \text{I} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \\ \times & \times & \times \\ 0 & \times & \times \\ \times & \times & \times \end{bmatrix} \end{array}.$$

Along the top row of the above identification matrix, "MP" stands for a shock to monetary policy, "P" stands for a shock to output production, and "I" stands for a shock to the information equation reflecting commodity markets. The symbol "×" means that no restriction is imposed on the long-run response, and "0" means an exclusion restriction. In this case, one can see that $n = 3$, $k = 2n = 6$, $q_1 = 2$, $q_2 = 1$, and $q_3 = 0$. The total number of restrictions $(\sum_{j=1}^{3} q_j)$ is equal to $n(n-1)/2 = 3$, and Rothenberg's (1971) order condition for exact identification holds. Because $q_j = n - j$ for $j = 1, \ldots, 3$, this model is exactly identified according to Theorem 7.

Now consider an alternative identification in which a shock in commodity markets has no long-run effect on output, but monetary policy has a long-run effect on output (i.e. we allow for a possible non-neutrality result). This alternative identification implies the following set of restrictions:

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{IR}_\infty \end{bmatrix} = \begin{array}{c} \\ \Delta \log P_c \\ \Delta \log Y \\ R \\ \Delta \log P_c \\ \Delta \log Y \\ R \end{array} \begin{array}{c} \begin{array}{ccc} \text{MP} & \text{P} & \text{I} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \\ \times & \times & \times \\ \times & \times & 0 \\ \times & \times & \times \end{bmatrix} \end{array}.$$

In this case, $n = 3$, $k = 2n = 6$, $q_1 = q_2 = q_3 = 1$. As in the case discussed in Section 4.4, one can show that the model is locally identified, but not identified (in the global sense) according to Theorem 7.

6. ALGORITHMS FOR ESTIMATION AND SMALL-SAMPLE INFERENCE

In Sections 2 and 4 we develop a general theory for global identification of SVARs. In Section 5, we apply our theory to a number of widely used models and show that slight modifications in the restrictions may render the model unidentifiable.

Once global identification has been established, the next step is to perform small-sample estimation and inference. For the maximum likelihood estimate or the posterior estimate when a prior is used, the existing estimation method for SVARs with restrictions on impulse responses

is inefficient. This inefficiency is more serious for small-sample inference because classical bootstrap procedures or Bayesian Monte Carlo Markov Chain (MCMC) methods require expensive random samples of the structural parameters (Kilian, 1998; Geweke, 1999; Inoue and Kilian, 2002; Geweke, 2005; Pesavento and Rossi, 2006). This problem becomes critical when time-varying SVARs are studied (Uhlig, 1997; Canova and Gambetti, 2004; Cogley and Sargent, 2005; Primiceri, 2005; Sims and Zha, 2006*b*; Gambetti and Canova, 2008). Consider an exactly identified model with drifting parameters and with restrictions directly imposed on impulse responses. The traditional method of Galí (1992) involves solving a system of nonlinear equations for every draw of the parameters at each time t. For a 12-lag model of more than three variables, for instance, using this method to obtain the number of draws often required to achieve accurate small-sample inferences quickly becomes computationally infeasible.

In what follows, we first build on Theorem 5 and develop an algorithm to achieve computational efficiency for exactly identified models. We show that for a triangular system, an even faster algorithm is feasible. Furthermore, we derive a computationally efficient algorithm designed for sign restrictions. This algorithm improves considerably on the existing methods and is important because sign restrictions have been widely used in the recent literature. Finally, we describe a class of priors that allow us to use these algorithms in the Bayesian framework.

6.1. *Algorithms for exactly identified models*

Assume that the model is exactly identified. Theorem 5 tells us that for almost any value of $(\mathbf{A}_0, \mathbf{A}_+)$, there is a unique orthogonal matrix $\mathbf{P}$ such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ satisfies the identifying restrictions. The matrix $\mathbf{P}$ is sometimes called the rotation matrix. The core of our argument is that, instead of solving a complicated system of nonlinear equations as in Galí (1992), we can find the rotation matrix $\mathbf{P}$ in a very efficient manner, as shown in the algorithm developed below. In this section we follow the convention that the transformation function $f(\cdot)$ has been so chosen that $\text{rank}(\mathbf{Q}_j) = q_j = n - j$ for $j = 1, \ldots, n$ as in Theorem 7.

Algorithm 1. *Consider an exactly identified SVAR with admissible and strongly regular restrictions represented by R. Let* $(\mathbf{A}_0, \mathbf{A}_+)$ *be any value of the unrestricted structural parameters.*

(Step 1) Set $j = 1$.

(Step 2) Form the matrix

$$\tilde{\mathbf{Q}}_j = \begin{bmatrix} \mathbf{Q}_j f(\mathbf{A}_0, \mathbf{A}_+) \\ \mathbf{p}_1' \\ \vdots \\ \mathbf{p}_{j-1}' \end{bmatrix}. \tag{10}$$

If $j = 1$, *then* $\tilde{\mathbf{Q}}_j = \mathbf{Q}_j f(\mathbf{A}_0, \mathbf{A}_+)$.

(Step 3) There exists a unit vector $\mathbf{p}_j$ *such that* $\tilde{\mathbf{Q}}_j \mathbf{p}_j = 0$ *because* $rank(\mathbf{Q}_j) = n - j$ *and hence* $rank(\tilde{\mathbf{Q}}_j) < n$. *Use the LU decomposition of* $\tilde{\mathbf{Q}}_j$ *to find this unit vector* $\mathbf{p}_j$. *The sign of* $\mathbf{p}_j$ *should be chosen to be consistent with the normalization rule.*

(Step 4) If $j = n$, *stop; otherwise, set* $j = j + 1$ *and go to Step 2.*

For any structural parameter point $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$, the above algorithm produces an orthogonal matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \ldots & \mathbf{p}_n \end{bmatrix},$$

such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R$. Theorem 5 guarantees that $\mathbf{P}$ will be unique for almost all structural parameters.

How does the algorithm work for small-sample estimation and inference? Suppose that one wishes to find the maximum likelihood (ML) estimate (or the estimate at the posterior peak) of the restricted model. Assume one is able to get the ML estimate (or the estimate at the posterior peak) for the unrestricted structural parameters or the reduced-form parameters.[12] Algorithm 1 provides us an orthogonal matrix $\mathbf{P}$ that rotates the unrestricted estimate to the estimate that satisfies the identifying restrictions. If the original estimate is for the reduced-form parameters, one can use Algorithm 1 to rotate the Cholesky decomposition of $\mathbf{\Sigma}$ to get the estimate of structural parameters that satisfy the restrictions.

Suppose now that one wishes to perform small-sample inference by using the bootstrap procedure or the Bayesian MCMC method to construct confidence intervals of structural parameters. Denote a draw of the unrestricted structural parameters by $(\mathbf{A}_0, \mathbf{A}_+)$. For such a draw, one uses Algorithm 1 to find the rotation matrix $\mathbf{P}$ such that $\mathbf{Q}_j f(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})\mathbf{e_j} = 0$ for all $1 \leq j \leq n$. If one makes a draw of the reduced-form parameters, one can obtain $(\mathbf{A}_0, \mathbf{A}_+)$ from the Cholesky decomposition of the reduced-form covariance matrix and proceed as just described.

6.2. *An example*

To illustrate how Algorithm 1 works in practice, we present a simple example of finding the rotation matrix $\mathbf{P}$ using Algorithm 1. For clarity of the exposition, we consider a simple three-variable standard SVAR with one lag so that $\mathbf{A}_+ = \mathbf{A}_1$, the analysis of which can be easily extended to more variables and more lags. The three variables are output growth ($\Delta \log Y$), the interest rate (R), and inflation ($\Delta \log P$). The three identifying restrictions are that demand shocks have no long-run effect on output, and that monetary policy shocks have neither a short-run nor a long-run effect on output. These restrictions on impulse responses can be expressed as the restrictions on the columns of the transformed matrix $f(\cdot)$. Table 3 presents this transformation and the restrictions on the transformed parameters. In the table, the symbol "$\times$" indicates no restrictions, "0" indicates an exclusion restriction, "MP" stands for monetary policy shocks, "D" stands for demand shocks, and "S" stands for supply shocks.

The first and foremost step is to determine whether this system is identified. From Table 3 one can see that $n = 3$, $q_1 = 2$, $q_2 = 1$, and $q_3 = 0$. It follows from Theorem 7 that this system is exactly identified. Therefore, for almost any value of $(\mathbf{A}_0, \mathbf{A}_+)$, there exists a unique rotation matrix $\mathbf{P}$ such that the restrictions hold.

TABLE 3
Short- and long-run restrictions

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{IR}_0 \\ \mathbf{IR}_\infty \end{bmatrix} = \begin{array}{c} \\ \Delta \log Y \\ R \\ \log P \\ \Delta \log Y \\ R \\ \log P \end{array} \begin{array}{c} \begin{array}{ccc} \text{MP} & \text{D} & \text{S} \end{array} \\ \begin{bmatrix} 0 & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \end{array}$$

12. Such an estimate of the unrestricted structural parameters may not be unique, but it gives the same likelihood or posterior value as other estimates at the peak of the likelihood or the posterior density.

To show how to find such a rotation matrix using Algorithm 1, we express the restrictions described by Table 3 in terms of the matrices $\mathbf{Q}_j$s as follows:

$$\mathbf{Q}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{and}$$

$$\mathbf{Q}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By deleting the rows of zeros from the above $\mathbf{Q}_j$s, we have

$$\overline{\mathbf{Q}}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \text{ and } \overline{\mathbf{Q}}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Since all the rows in $\mathbf{Q}_3$ are zeros, there is no $\overline{\mathbf{Q}}_3$. Working with $\overline{\mathbf{Q}}_j$ is operationally easier than working with $\mathbf{Q}_j$, since $\tilde{\mathbf{Q}}_j$ in Algorithm 1 derived from $\overline{\mathbf{Q}}_j$ will always be an $(n-1) \times n$ matrix.[13]

For the purpose of walking through Algorithm 1, suppose that the estimates of the reduced-form parameters $\mathbf{B}$ and $\mathbf{\Sigma}$ are

$$\mathbf{B} = \begin{bmatrix} 0.5 & -1.25 & -1 \\ 0.5 & 0.25 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.5 & 1 \\ 0.5 & 4.25 & 2.5 \\ 1 & 2.5 & 3 \end{bmatrix}.$$

We compute $\mathbf{A}_0$ from the Cholesky decomposition of $\mathbf{\Sigma}^{-1}$ (in Matlab, $\mathbf{A}_0' = \text{chol}(\mathbf{\Sigma}^{-1})$), and $\mathbf{A}_+ = \mathbf{B}\mathbf{A}_0$. Since $\mathbf{IR}_0' = \mathbf{A}_0^{-1}$ and $\mathbf{IR}_\infty' = (\mathbf{A}_0 - \mathbf{A}_1)^{-1}$, we have

$$f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} \mathbf{IR}_0 \\ \mathbf{IR}_\infty \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{11}$$

and

$$\tilde{\mathbf{Q}}_1 = \overline{\mathbf{Q}}_1 f(\mathbf{A}_0, \mathbf{A}_+) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

13. Since the only difference between $\mathbf{Q}_j$ and $\overline{\mathbf{Q}}_j$ is rows of zeros, $\mathbf{Q}_j f(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})\mathbf{e_j} = 0$ and $\overline{\mathbf{Q}}_j f(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})\mathbf{e_j} = 0$ are equivalent statements. In many applications, it is obvious how to form $\overline{\mathbf{Q}}_j$, but one can always use the Matlab function orth() and define $\overline{\mathbf{Q}}_j' = \text{orth}(\mathbf{Q}_j')$.

The first step in Algorithm 1 is to find a unit length vector $\mathbf{p}_1$ such that $\tilde{\mathbf{Q}}_1\mathbf{p}_1 = 0$. The most computationally efficient method of finding this vector is to employ the LU decomposition of $\tilde{\mathbf{Q}}_1$. However, employing the QR decomposition of $\tilde{\mathbf{Q}}_1'$ is often more convenient.[14] Let $\tilde{\mathbf{Q}}_1' = \mathbf{QT}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{T}$ is upper triangular. If we choose $\mathbf{p}_1$ to be the last row of $\mathbf{Q}$, then $\tilde{\mathbf{Q}}_1\mathbf{p}_1$ will be the last column of $\mathbf{T}'$, which is zero. Therefore, in this example $\mathbf{p}_1' = [0 \quad 0 \quad 1]$.

To obtain $\mathbf{p}_2$, we form

$$\tilde{\mathbf{Q}}_2 = \begin{bmatrix} \overline{\mathbf{Q}}_2 f(\mathbf{A}_0, \mathbf{A}_+) \\ \mathbf{p}_1' \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

As before, take $\mathbf{p}_2$ to be the last row of the orthogonal component of the QR decomposition of $\tilde{\mathbf{Q}}_2'$ to get $\mathbf{p}_2' = [-0.7071 \quad 0.7071 \quad 0]$.

To obtain $\mathbf{p}_3$, we form

$$\tilde{\mathbf{Q}}_3 = \begin{bmatrix} \mathbf{p}_1' \\ \mathbf{p}_2' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0.7071 & -0.7071 & 0 \end{bmatrix}.$$

Again, letting $\mathbf{p}_3$ be the last row of the orthogonal component of the QR decomposition of $\tilde{\mathbf{Q}}_3'$, we get $\mathbf{p}_3' = [-0.7071 \quad -0.7071 \quad 0]$.

Combining these results, one obtains the orthogonal matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix} = \begin{bmatrix} 0 & -0.7071 & -0.7071 \\ 0 & 0.7071 & -0.7071 \\ 1 & 0 & 0 \end{bmatrix}.$$

Verifying that $\mathbf{Q}_j f(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})\mathbf{e_j} = 0$ for $1 \leq j \leq 3$ is a simple process.

6.3. *Algorithms for triangular systems*

While Algorithm 1 gives us an efficient way to find the rotation matrix $\mathbf{P}$ for exactly identified models, in this section we present a much faster algorithm for triangular systems, which are defined as follows:

Definition 6. *Consider an SVAR with admissible restrictions represented by R. The system is triangular if and only if there exists a $n \times n$ permutation matrix $\mathbf{P}_0$ and an invertible $k \times k$ matrix $\mathbf{P}_1$ such that the structural parameter point $(\mathbf{A}_0, \mathbf{A}_+)$ satisfies the restriction if and only if $\mathbf{P}_1 f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{P}_0$ is lower triangular.*

This definition is equivalent to the one discussed in Section 4 in the context of linear restrictions on the contemporaneous coefficient matrix $\mathbf{A}_0$. If the columns of $(\mathbf{A}_0, \mathbf{A}_+)$ have been ordered according to the convention given by (5), then $\mathbf{P}_0$ will be the matrix with ones along the anti-diagonal. Multiplication on the right by this matrix has the effect of reversing the ordering of columns. For a triangular system, Algorithm 1 can be so improved that the

14. In Matlab, the function $qr(\cdot)$ applied to an $n \times (n-1)$ matrix returns an $n \times n$ orthogonal matrix $\mathbf{Q}$ and an $n \times (n-1)$ upper triangular matrix R. In other software packages, however, the "orthogonal" matrix $\mathbf{Q}$ may be $n \times (n-1)$ and the triangular matrix R may be $(n-1) \times (n-1)$. If those packages are used, one needs to pad the matrix $\tilde{\mathbf{Q}}_j$ with a row of zeros before proceeding further. In either case, the last row of R will be zero.

orthogonal matrix given by Theorem 5 can be found using only a single (instead of successive) QR decomposition as described in the following theorem.

Theorem 8. *Consider an exactly identified triangular SVAR with admissible and strongly regular restrictions represented by R. Let $\mathbf{P}_0$ and $\mathbf{P}_1$ be the matrices that make the restrictions triangular and let $(\mathbf{A}_0, \mathbf{A}_+)$ be any structural parameter point. If $(\mathbf{P}_1 f(\mathbf{A}_0, \mathbf{A}_+))' = \mathbf{QT}$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{T}$ is upper triangular, then $\mathbf{P} = \mathbf{QP}_0'$ is the orthogonal matrix such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R$.*

Proof. Since $\mathbf{P}_1 f(\mathbf{A}_0, \mathbf{A}_+) = \mathbf{T}'\mathbf{Q}'$,

$$\begin{aligned}\mathbf{T}' &= \mathbf{P}_1 f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{QP}_0'\mathbf{P}_0 \\ &= \mathbf{P}_1 f(\mathbf{A}_0\mathbf{QP}_0', \mathbf{A}_+\mathbf{QP}_0')\mathbf{P}_0.\end{aligned}$$

Since $\mathbf{T}'$ is lower triangular, $(\mathbf{A}_0\mathbf{QP}_0', \mathbf{A}_+\mathbf{QP}_0')$ must satisfy the restrictions. ||

In the example of Section 6.2, the restrictions are exclusion restrictions, i.e. certain parameters are set to zero. For exclusion restrictions, the system will be triangular if and only if permutations of the rows and columns make the system lower triangular. If we define

$$\mathbf{P}_0 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{P}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

then $\mathbf{P}_0$ and $\mathbf{P}_1$ permute the rows and columns of the example of Section 6.2 into a triangular system. Thus, we can apply Theorem 8 to find the rotation matrix $\mathbf{P}$. As an illustration, we use the same values of the reduced-form parameters $\mathbf{B}$ and $\mathbf{\Sigma}$ as in Section 6.2. The numerical value of the matrix $f(\mathbf{A}_0, \mathbf{A}_+)$ is given by (11). The QR decomposition of $(\mathbf{P}_1 f(\mathbf{A}_0, \mathbf{A}_+))'$ gives

$$\mathbf{Q} = \begin{bmatrix} -0.7071 & -0.7071 & 0 \\ -0.7071 & 0.7071 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The required rotation matrix is

$$\mathbf{P} = \mathbf{QP}_0 = \begin{bmatrix} 0 & -0.7071 & -0.7071 \\ 0 & 0.7071 & -0.7071 \\ 1 & 0 & 0 \end{bmatrix},$$

which is identical to the $\mathbf{P}$ obtained in Section 6.2.

6.4. *Sign restrictions*

The identifying restrictions described in Section 2.3 and the algorithms developed in Sections 6.1 and 6.3 are based on equality restrictions on the transformed structural parameters. One objective in studying this class of restrictions is to identify structural shocks. For example, according to the conventional wisdom (and many DSGE models), a contractionary monetary

policy shock should raise the interest rate and lower output and prices. Thus, a successful identification should produce impulse responses that conform to this conventional wisdom. Some identification schemes as described in Section 2.3 may not generate impulse responses that have the desired signs. In this situation, Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) propose an alternative approach. Their basic idea is to use sign restrictions imposed directly on impulse responses. For example, the interest rate rises while money, output, and prices fall in response to a contractionary monetary shock.

Unfortunately, the algorithms established in Sections 6.1 and 6.3 cannot be applied to small-sample estimation and inference of an SVAR with sign restrictions because an SVAR with sign restrictions on impulse responses is not even locally identified. For any set of sign restrictions, given a parameter point $(\mathbf{A}_0, \mathbf{A}_+)$ that satisfies such restrictions, there always exists an orthogonal matrix $\mathbf{P}$, arbitrarily close to an identity matrix, such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ will also satisfy the sign restrictions.[15] Given the fact that an SVAR with only sign restrictions is not identified, one important task is to find a set of impulse responses that satisfy the same sign restrictions (Fry and Pagan, 2007). This task is computationally intensive and amounts to finding a set of $\mathbf{P}$s such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ satisfies the sign restrictions.

Canova and DeNicoló (2002) propose an algorithm based on grid search to find such a $\mathbf{P}$. This algorithm, however, becomes computationally infeasible when an SVAR system is moderately large (e.g. $n > 4$). To solve this inefficiency problem, we develop a new algorithm, based on the following theorem:

Theorem 9. *Let $\tilde{\mathbf{X}}$ be an $n \times n$ random matrix with each element having an independent standard normal distribution. Let $\tilde{\mathbf{X}} = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$ be the QR decomposition of $\tilde{\mathbf{X}}$ with the diagonal of $\tilde{\mathbf{R}}$ normalized to be positive. Then $\tilde{\mathbf{Q}}$ has the uniform (or Haar) distribution.*

Proof. The proof follows directly from Stewart (1980).[16] ||

Theorem 9 gives us a convenient way of implementing a random selection of orthogonal matrices to obtain impulse responses that satisfy the sign restrictions. The following algorithm describes this implementation.

Algorithm 2. *Let $(\mathbf{A}_0, \mathbf{A}_+)$ be any given value of the unrestricted structural parameters.*
(Step 1) Draw an independent standard normal $n \times n$ matrix $\tilde{\mathbf{X}}$ and let $\tilde{\mathbf{X}} = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$ be the QR decomposition of $\tilde{\mathbf{X}}$ with the diagonal of $\tilde{\mathbf{R}}$ normalized to be positive.
(Step 2) Let $\mathbf{P} = \tilde{\mathbf{Q}}$ and generate impulse responses from $\mathbf{A}_0\mathbf{P}$ and $\mathbf{B} = \mathbf{A}_+\mathbf{A}_0^{-1}$.
(Step 3) If these impulse responses do not satisfy the sign restrictions, return to Step 1.

If $(\mathbf{A}_0, \mathbf{A}_+)$ is a draw of unrestricted parameters, $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ obtained via Algorithm 2 is a draw of structural parameters satisfying the sign restrictions.[17] If $\mathbf{B}$ and $\mathbf{\Sigma}$ are a draw of the reduced-form parameters, we use $\mathbf{B}$ and the Cholesky decomposition of $\mathbf{\Sigma}$ to obtain $(\mathbf{A}_0, \mathbf{A}_+)$ and then use Algorithm 2 to obtain $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ that satisfies the sign restrictions.

15. This point is not new and has been made by, for example, Geweke (1986) in the context of inequality restrictions in linear regression models.

16. Stewart (1980) has even more efficient algorithms for generating uniform random orthogonal matrices, but they are less straightforward and more difficult to implement.

17. In theory the algorithm is not guaranteed to terminate. In practice, we set a maximum number of iterations to be 100,000 for Steps 2–3 to be repeated. If the maximum is reached, the algorithm should move to Step (1) to draw another orthogonal matrix $\tilde{\mathbf{Q}}$.

Algorithm 2 describes how to randomly search for a rotation matrix $\mathbf{P}$ for a given value of $(\mathbf{A}_0, \mathbf{A}_+)$ or $(\mathbf{B}, \boldsymbol{\Sigma})$. If one repeats Steps 1–3 to generate draws of $\mathbf{P}$ for the given value of $(\mathbf{A}_0, \mathbf{A}_+)$ or $(\mathbf{B}, \boldsymbol{\Sigma})$, one effectively obtains a set of $\mathbf{P}$s such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ satisfies the sign restrictions.

An even greater efficiency gain can be obtained by realizing that changing the sign of any column of the obtained orthogonal matrix $\mathbf{P}$ results in another orthogonal matrix. In Step 3, if all the impulse responses generated by $\mathbf{A}_0\mathbf{P}$ and $\mathbf{B}$ have wrong signs, one could simply change the sign of the j-th column of $\mathbf{P}$. The impulse responses generated with this new rotation matrix would satisfy the sign restrictions. This sign change would increase the algorithm's efficiency by a factor of 2^n.[18]

Our algorithm should be viewed as a generalization of Uhlig's (2005) algorithm. If several structural shocks must be identified, Uhlig's (2005) algorithm searches for the orthogonal matrix column by column recursively. During this search, the orthogonal matrix may not be found for some draws of $(\mathbf{B}, \boldsymbol{\Sigma})$. Our algorithm, based on the Householder-transformation methodology, differs from Uhlig's (2005) algorithm in two aspects: (1) all the draws of $(\mathbf{B}, \boldsymbol{\Sigma})$ are kept in practice, and (2) for a given value of $(\mathbf{B}, \boldsymbol{\Sigma})$, the orthogonal matrix is a random draw from the uniform (or Haar) distribution with only a single operation of the QR decomposition. These differences make our algorithm more efficient when several shocks are to be identified. In particular, for a time-varying SVAR system with more than three or four structural shocks to be identified, the efficiency gain can be as high as the 10-1 ratio when our algorithm is compared with the algorithms of Uhlig (2005) and Canova and DeNicoló (2002) (see, for example, Benati and Mumtaz (2006) for an application of our algorithm in their DSGE model).[19]

6.5. *A reference prior*

Sections 6.1, 6.3, and 6.4 develop new algorithms usable for small-sample estimation and inference in a large class of SVARs. In the Bayesian framework, an additional property for the prior is needed. If $\mathbf{P}$ is an orthogonal matrix, the transformed parameters $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ must have the same prior density as the original parameters $(\mathbf{A}_0, \mathbf{A}_+)$. This property ensures that our algorithms are valid for Bayesian estimation and inference as well as for computation of the marginal data density for model comparison.[20]

In this subsection, we show that the reference prior of Sims and Zha (1998) is such that $(\mathbf{A}_0, \mathbf{A}_+)$ and $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$ have the same prior density. The Sims and Zha prior is general and encompasses the popular Minnesota prior. Let $\boldsymbol{a}_0 = \text{vec}(\mathbf{A}_0)$ and $\boldsymbol{a}_+ = \text{vec}(\mathbf{A}_+)$. The prior distribution for the unrestricted parameters takes the following form:

$$\boldsymbol{a}_0 \sim N(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{H}_0), \tag{12}$$

and

$$\boldsymbol{a}_+ \mid \boldsymbol{a}_0 \sim N(\text{vec}(\overline{\mathbf{S}}\mathbf{A}_0), \mathbf{I}_n \otimes \mathbf{H}_+), \tag{13}$$

where $\mathbf{H}_0$ and $\mathbf{H}_+$ are symmetric positive definite matrices and $\overline{\mathbf{S}}$ is typically set as

$$\begin{bmatrix} \mathbf{I}_{n\times n} \\ \mathbf{0}_{m\times n} \end{bmatrix}$$

to represent a random-walk component of the prior.

18. We thank an anonymous referee for drawing our attention to this point.

19. Note that since the particular application studied by Uhlig (2005) concerns a monetary policy shock and not any other shocks, the efficiency gain from our algorithm is negligible in this application.

20. We thank an anonymous referee for bringing out this important point.

Because $\text{vec}(\mathbf{A}_0\mathbf{P}) = (\mathbf{P}' \otimes \mathbf{I}_n)\mathbf{A}_0$, $\text{vec}(\bar{\mathbf{S}}\mathbf{A}_0\mathbf{P}) = (\mathbf{P}' \otimes \mathbf{I}_n)\text{vec}(\bar{\mathbf{S}}\mathbf{A}_0)$, $\text{vec}(\mathbf{A}_+\mathbf{P}) = (\mathbf{P}' \otimes \mathbf{I}_n)\mathbf{A}_+$, and $\mathbf{P}' \otimes \mathbf{I}_n$ commutes with both $\mathbf{I}_n \otimes \mathbf{H}_0$ and $\mathbf{I}_n \otimes \mathbf{H}_+$, if $\mathbf{P}$ is orthogonal, then (12) and (13) imply that the prior density of $(\mathbf{A}_0, \mathbf{A}_+)$ is equal to the prior density of $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$. Since an orthogonal rotation of $(\mathbf{A}_0, \mathbf{A}_+)$ does not change the likelihood, it is also the case that the posterior density of $(\mathbf{A}_0, \mathbf{A}_+)$ is equal to the posterior density of $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$.

The proper prior we have discussed thus far concerns the unrestricted structural parameters. As shown in Geweke (1986), if an unrestricted model has a proper prior, then after imposing identifying restrictions the prior is still proper. The posterior for the restricted model will be proper as well.

7. CONCLUSION

SVARs are widely used for policy analysis and to provide stylized facts for economic theory. Before one performs an empirical analysis with a particular SVAR, however, analysing the model to ascertain its identifiability is essential. Otherwise, the empirical results would be misleading.

In this paper, we establish general rank conditions for global identification of both identified and exactly identified models. These conditions are sufficient for identification and necessary and sufficient for exact identification; they can be checked via a simple matrix-filling and rank-checking exercise.

The usual sufficient rank conditions for (local) identification in the literature involve computing the rank of certain derivatives (Fisher, 1966; Rothenberg, 1971; Hausman and Taylor, 1983). Because we are able to obtain much simpler conditions by exploiting the orthogonal structure of the problem, we are not required to explicitly compute any derivatives. For exactly identified SVARs, our necessary and sufficient condition involves counting the number of restrictions and checking their pattern, an exercise no more complicated than Rothenberg's (1971) order condition.

Finally, the efficient algorithms developed in this paper provide essential tools for researchers studying a variety of empirical SVARs.

APPENDIX A. PROOFS OF THEOREMS 3 AND 5

Throughout these appendices, we will need conditions that ensure that the image or the preimage of a set of measure zero will be of measure zero. The following two lemmas will suffice for our purposes. These lemmas follow from the inverse function theorem, Sard's theorem, and the implicit function theorem. A statement of all of these can be found in Spivak (1965).

Lemma 1. *Let M_1 and M_2 be differentiable manifolds of the same dimension and let h be a continuously differentiable function from M_1 to M_2. If E is a set of measure zero in M_1, then $h(E)$ is of measure zero in M_2.*

Lemma 2. *Let M_1 be an i-dimensional differentiable manifold and M_2 a j-dimensional differentiable manifold and let h be a continuously differentiable function from M_1 to M_2 whose derivative is of rank j for all $x \in M_1$. If $\hat{M}_2$ is a differentiable submanifold of M_2 and E is a set of measure zero in $\hat{M}_2$, then $\hat{M}_1 = f^{-1}(\hat{M}_2)$ is a differentiable submanifold of M_1 and $h^{-1}(E)$ is of measure zero in $\hat{M}_1$.*

Theorem 5 follows directly from Lemma 1.

Proof of Theorem 5. Consider the function h from $\mathbb{P}^R \times O(n)$ to $\mathbb{P}^S$ to which maps $(\mathbf{B}, \mathbf{\Sigma}) \times \mathbf{P}$ to $(\mathbf{TP}, \mathbf{BTP})$, where $\mathbf{T}$ is the unique lower triangular matrix with positive diagonal such that $\mathbf{TT}' = \mathbf{\Sigma}^{-1}$. The matrix $\mathbf{T}$ can be

obtained from the Cholesky decomposition of $\mathbf{\Sigma}^{-1}$. The function h is continuously differentiable with continuously differentiable inverse. Define $\hat{G}$ by

$$\left\{(\mathbf{B}, \mathbf{\Sigma}) \in \mathbb{P}^R \mid \text{there is not a unique } (\mathbf{A}_0, \mathbf{A}_+) \in R \text{ s.t. } g(\mathbf{A}_0, \mathbf{A}_+) = (\mathbf{B}, \mathbf{\Sigma})\right\} \tag{14}$$

and G by

$$\{(\mathbf{A}_0, \mathbf{A}_+) \in U \mid \text{there is not a unique } \mathbf{P} \in O(n) \text{ s.t. } (\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R\}. \tag{15}$$

Definition 4.1 states that an SVAR model is exactly identified if and only if $\hat{G}$ is of measure zero and Theorem 5 states that an SVAR is exactly identified if and only if G is of measure zero. Since $G = h(\hat{G} \times O(n))$ and $\hat{G} \times O(n) = h^{-1}(G)$, Lemma 1 gives that G is of measure zero if and only if $\hat{G} \times O(n)$ is of measure zero. Theorem 5 now follows from the fact that $\hat{G} \times O(n)$ is of measure zero if and only if $\hat{G}$ is of measure zero. ||

Before proving Theorem 3, we develop more notation and prove an additional lemma. The notation will be used through out these appendices. Because we allow quite general normalization rules, R may not be a differentiable manifold. However, the set of un-normalized restricted structural parameters will be. We denote this set by $\mathcal{R}$ and

$$\mathcal{R} = \left\{(\mathbf{A}_0, \mathbf{A}_+) \in U \mid \mathbf{Q}_j f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{e}_\mathbf{j} = 0 \text{ for } 1 \leq j \leq n\right\}. \tag{16}$$

The set R will be equal to $\mathcal{R} \cap N$. Related to $\mathcal{R}$ is the set $\tilde{\mathcal{R}}$ defined by

$$\tilde{\mathcal{R}} = \left\{\mathbf{X} \in f(U) \mid \mathbf{Q}_j\mathbf{X}\mathbf{e}_\mathbf{j} = 0 \text{ for } 1 \leq j \leq n\right\}. \tag{17}$$

The set $\tilde{\mathcal{R}}$ is a open subset of a linear subspace of the set of $k \times n$ matrices, $f(\mathcal{R}) = \tilde{\mathcal{R}}$, and $f^{-1}(\tilde{\mathcal{R}}) = \mathcal{R}$. The dimension of $\tilde{\mathcal{R}}$ is $nk - q$. Finally, define $\tilde{R}$ by

$$\tilde{R} = f(R). \tag{18}$$

When we say, as was done in Theorem 3, that a subset E of R is of measure zero in R, we take this to mean that E is of measure zero in the manifold $\mathcal{R}$.

The following lemma will be a key ingredient in the proof of Theorem 3.

Lemma 3. *For $1 \leq j \leq n$, let V_j be a linear subspace of $\mathbb{R}^m$ and let $V = V_1 \times \cdots \times V_n$. Define S to be the set of all $(\mathbf{v}_1, \ldots, \mathbf{v}_n) \in V$, whose span is of dimension strictly less than n. Either $S = V$ or S is a set of measure zero in V.*

Proof. Let χ_S be the indicator function of S. To prove that S is of measure zero, it suffices to show that $\int_V \chi_S = 0$. We divide $V_1 \times \cdots \times V_{n-1}$ into two sets A and B. We will show that A is of measure zero and that $\int_{V_n} \chi_S(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) = 0$ for every $(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) \in B$. This implies that

$$\begin{aligned}\int_V \chi_S &= \int_{V_1 \times \cdots \times V_{n-1}} \int_{V_n} \chi_S(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) \\ &= \int_A \int_{V_n} \chi_S(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) + \int_B \int_{V_n} \chi_S(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) = 0\end{aligned}$$

as desired.

We proceed by induction on n. When $n = 1$, $V = V_1$, and $S = \{0\}$. If the dimension of V is zero, then $S = V$, and if the dimension of V is positive, then S is of measure zero in V.

Now assume that the lemma holds for $n - 1$. Suppose that $S \neq V$, so there exists $(\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n) \in V$, whose span is of dimension n. Define $\tilde{V}_n$ to be the $m - 1$ dimensional subspace of $\mathbb{R}^m$ that is perpendicular $\tilde{v}_n$ and let ρ be the projection mapping onto $\tilde{V}_n$. For $1 \leq j \leq n - 1$, define $W_j = \rho(V_j)$ and let $W = W_1 \times \cdots \times W_{n-1}$. Define S_{n-1} to be the elements of W whose span is of dimension less than $n - 1$. By the induction hypothesis, either $S_{n-1} = W$ or S_{n-1} is of measure zero in W. Since the span of $\rho(\tilde{\mathbf{v}}_1), \ldots, \rho(\tilde{\mathbf{v}}_{n-1})$ is of dimension $n - 1$, $(\rho(\tilde{\mathbf{v}}_1), \ldots, \rho(\tilde{\mathbf{v}}_{n-1})) \notin S_{n-1}$ and $S_{n-1} \neq W$. Thus, S_{n-1} is of measure zero in W. Let A be the set of all $(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) \in V_1 \times \cdots \times V_{n-1}$ such that $(\rho(\mathbf{v}_1), \ldots, \rho(\mathbf{v}_{n-1})) \in S_{n-1}$. Because S_{n-1} is of measure zero in W, A must be of measure zero in V, as desired. All that remains to be shown is that $\int_{V_n} \chi_S(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) = 0$ for every $(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) \in B$, where B is the complement of A in $V_1 \times \cdots \times V_{n-1}$.

Fix $(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) \in B$ and let C be the span of $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$. The dimension of C is $n-1$ because $(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) \in B$, and hence $\chi_S(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) \neq 0$ if and only if $\mathbf{v}_n \in C$. Thus, $\int_{V_n} \chi_S(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) = 0$ if and only if $V_n \cap C$ is of measure zero in V_n. Since $V_n \cap C$ is either all of V_n or a set of measure zero in V_n, it suffices to show that there exists an element of V_n that is not in C. Again, since $(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) \in B$, the vector $\tilde{v}_n$ is not in C. ||

We are now ready to complete the proof of Theorem 3.

Proof of Theorem 3. Let W be the complement of K in R. In terms of W, Theorem 3 states that either $W = R$ or W is of measure zero in $\mathcal{R}$. Let $\tilde{W} = f(W)$. From the definition of K given by (7) one can easily check that $f^{-1}(\tilde{W}) = W$. So, by Lemma 2, we have that if $\tilde{W}$ is of measure zero in $\tilde{\mathcal{R}}$, then W is of measure zero in $\mathcal{R}$. Thus, if suffices to prove that either $\tilde{W} = \tilde{R}$ or $\tilde{W}$ is of measure zero in $\tilde{\mathcal{R}}$.

Let $\tilde{V}$ be the set of all $k \times n$ matrices $\mathbf{X}$ such that $\mathbf{Q}_j \mathbf{X} \mathbf{e_j} = 0$ for $1 \leq j \leq n$, and let $\tilde{W}_i$ be the set of all matrices $\mathbf{X} \in \tilde{V}$ such that $\text{rank}(\mathbf{M}_i(\mathbf{X})) < n$. Since $\tilde{R} \subset \tilde{V}$ and $\tilde{W} = \bigcap_{i=1}^{n} \tilde{W}_i \cap \tilde{R}$, it suffices to show that either $\tilde{W}_i = \tilde{V}$ or $\tilde{W}_i$ is of measure zero in $\tilde{V}$. Because of the block structure of $\mathbf{M}_i(\cdot)$, if $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x_n}]$, then $\mathbf{X} \in \tilde{W}_i$ if and only if $\text{rank}(\mathbf{Q}_i[\mathbf{x_{i+1}}, \ldots, \mathbf{x_n}]) < n - i$. The results now follow from Lemma 3 with V_j defined to be the set of all vectors of the form $\mathbf{Q}_i \mathbf{x}$ where $\mathbf{Q}_{i+j} \mathbf{x} = 0$. ||

From the proof of Theorem 3, it easily follows that if

$$K_j = \left\{ (\mathbf{A}_0, \mathbf{A}_+) \in R \mid \text{rank}\left(\mathbf{M}_j\left(f\left(\mathbf{A}_0, \mathbf{A}_+\right)\right)\right) = n \right\}, \tag{19}$$

then either K_j is empty or the complement of K_j in R is of measure zero in R. We record this in the following lemma that will be used in Appendix B.

Lemma 4. *Either K_j is empty or the complement of K_j in R is of measure zero in R.*

APPENDIX B. PROOFS OF THEOREMS 6 AND 7

We prove Theorem 7 first, and then Theorem 6. We proceed via a sequence of lemmas.

Lemma 5. *If $q_j = n - j$ for $1 \leq j \leq n$, then for every $(\mathbf{A}_0, \mathbf{A}_+) \in U$, there exists a $\mathbf{P} \in O(n)$ such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R$.*

Proof. Algorithm 1 explicitly shows how to construct such an orthogonal matrix $\mathbf{P}$ for any structural parameter point $(\mathbf{A}_0, \mathbf{A}_+)$. ||

Lemma 6. *If $q_j = n - j$ for $1 \leq j \leq n$, then there exists $(\mathbf{A}_0, \mathbf{A}_+) \in R$ such that $\mathbf{M}_j(\mathbf{A}_0, \mathbf{A}_+)$ is of rank n for $1 \leq j \leq n$.*

Proof. In light of Lemma 4, it suffices to construct a matrix $\mathbf{X}_i \in \tilde{\mathcal{R}}$ such that $\mathbf{M}_i(\mathbf{X}_i)$ is of rank n. It follows from Lemma 5 that $\tilde{\mathcal{R}}$ is non-empty, so let $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$ be any matrix in $\tilde{\mathcal{R}}$. Let $\mathbf{Y}_i^j = [\mathbf{y}_i, \ldots, \mathbf{y}_j]$. The matrix $\mathbf{M}_i(\mathbf{Y})$ is of rank n if and only if the matrix $\mathbf{Q}_i \mathbf{Y}_{i+1}^n$ is of rank $n - i$. Let V_j be the column space of $\mathbf{Q}_i \mathbf{Y}_{i+1}^{i+j}$. If the dimension of V_j is j for $1 \leq j \leq n - i$, then $\mathbf{M}_i(\mathbf{Y})$ is of rank n. If this is not the case, let j be the first index such that the dimension of V_j is less than j. Because the dimension of V_{j-1} is $j - 1$ and the dimension of the null space of $\mathbf{Q}_i$ is i, the dimension of the set of all vectors $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{Q}_i \mathbf{v} \in V_{j-1}$ is at most $i + j - 1$. Since the dimension of the null space of $\mathbf{Q}_{i+j}$ is $i + j$, there is a non-zero vector $\mathbf{v}$ in the null space of $\mathbf{Q}_{i+j}$ such that $\mathbf{Q}_i \mathbf{v}$ is not in V_{j-1}. Because $f(U)$ is an open set and $\mathbf{Y} \in \tilde{\mathcal{R}}$, there exists $\varepsilon > 0$ such that if we replace $\mathbf{y}_{i+j}$ by $\mathbf{y}_{i+j} + \varepsilon \mathbf{v}$, then the resulting matrix will be an element of $\tilde{\mathcal{R}}$ and have the property that the dimension of the column space of $\mathbf{Q}_i[\mathbf{y}_{i+1}, \ldots, \mathbf{y}_{i+j-1}, \mathbf{y}_{i+j} + \varepsilon \mathbf{v}]$ will be of dimension j. So, starting with any $\mathbf{Y} \in \tilde{\mathcal{R}}$, we can sequentially modify $\mathbf{Y}$ until we arrive at a matrix $\mathbf{x}_i \in \tilde{\mathcal{R}}$ such that $\mathbf{M}_i(\mathbf{X}_i)$ is of rank n. ||

Given these two lemmas, the following is now easy to show.

Lemma 7. *If $q_j = n - j$ for $1 \leq j \leq n$, then the SVAR model is exactly identified.*

Proof. Let G be defined by (15), K by (7), and W by

$$W = \{(\mathbf{A}_0, \mathbf{A}_+) \in R \mid \text{there exists } \mathbf{P} \in O(n) \text{ s.t. } \mathbf{P} \neq \mathbf{I}_n \text{ and } (\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R\}$$

We must show that G is of measure zero. By Theorem 1, W is contained in the complement of K, and so by Lemma 6 and Theorem 3, W is of measure zero in $\mathcal{R}$. Let $h : \mathcal{R} \times O(n) \to U$ be the continuously differentiable function which maps $(\mathbf{A}_0, \mathbf{A}_+) \times \mathbf{P}$ to $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P})$. It follows from Lemma 5 that $h(W \times O(n)) = G$. Since $W \times O(n)$ is of measure zero, by Lemma 1 the set $\mathbf{G}$ will be of measure zero. ||

One direction in the proof of Theorem 7 is now complete. The following lemma is key to proving the other direction,

Lemma 8. *Let $i \leq n \leq k$. If $V_1, \ldots, V_i$ are subspaces of $\mathbb{R}^k$ with the dimension of V_j equal to d_j for $1 \leq j \leq i$ and for all $k \times n$ matrices $\mathbf{X}$ there exist orthonormal vectors $\mathbf{p}_1, \ldots, \mathbf{p}_i$ in $\mathbb{R}^n$ such that $\mathbf{X}\mathbf{p}_j \in V_j$ for $1 \leq j \leq i$, then there exists a j such that $1 \leq j \leq i$ and $d_j \geq k - n + i$.*

Before giving the formal proof of this lemma, we explain the geometry behind the result. Consider the case $i = n = k = 3$. Since the implication of the lemma is that at least one of the subspaces must be all of $\mathbb{R}^3$, suppose that each of subspaces V_j is at most a plane. It is easy to see that there exists a line L in $\mathbb{R}^3$ that intersects each V_j only at the origin. Let K be the plane through the origin that is perpendicular to L, and let $\mathbf{Y}$ be the 3×3 matrix which projects $\mathbb{R}^3$ onto K along L. While $\mathbf{Y}$ is not invertible, there are invertible matrices that are arbitrarily close to $\mathbf{Y}$. If $\mathbf{X}$ is an invertible matrix such that $\mathbf{X}^{-1}$ is close to $\mathbf{Y}$, then the subspace $\mathbf{X}^{-1}V_j$ will almost lie in K. This result depends crucially on the fact that L intersects V_j only at the origin. Since three orthonormal vectors cannot possibly lie almost in a plane, it cannot be the case that for all 3×3 matrices $\mathbf{X}$ there exist orthonormal vectors $\mathbf{p}_j$ such that $\mathbf{p}_i \in \mathbf{X}^{-1}V_j$. Thus, the theorem holds for $i = n = k = 3$. The formal proof simply makes rigorous what is geometrically intuitive in this special case.

Proof. The proof will use the following three facts:

1. For $0 \leq d \leq k$, there exists a subspace U of $\mathbb{R}^k$ of dimension d such that the dimension of of $V_j \cap U$ is equal to $\max\{0, d + d_j - k\}$ for $1 \leq j \leq i$.
2. Let W be a $i - 1$ dimensional subspace of $\mathbb{R}^n$. There exists a $\delta > 0$ such that there cannot be i orthonormal vectors in the set
$$S_{W,\delta} = \left\{\mathbf{w} + \mathbf{u} \in \mathbb{R}^k \mid \mathbf{w} \in W \text{ and } \|\mathbf{u}\| < \delta\right\}.$$
3. Let U and V be subspaces of $\mathbb{R}^n$ such that $U \cap V = \{0\}$ and let W be the perpendicular complement of U. For every $\varepsilon > 0$, let $\mathbf{X}_{W,\varepsilon}$ be the changed to matrix that fixes W and maps each $\mathbf{u} \in U$ to $\varepsilon\mathbf{u}$. For every $\delta > 0$ there exists a $\gamma > 0$ such that for all $\gamma > \varepsilon > 0$ if $\mathbf{X}_{W,\varepsilon}^{-1}\mathbf{v} \in V$ and $\|\mathbf{v}\| = 1$, then $\mathbf{v} \in S_{W,\delta}$.

Using (1), we see that there exists a subspace U of $\mathbb{R}^k$ of dimension n such that the dimension of $V_j \cap U$ is of dimension $\tilde{d}_j = \max\{0, n + d_j - k\}$. Let $\mathbf{X}$ be a $k \times n$ matrix whose column space if equal to U. Let

$$\tilde{V}_j = \left\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{X}\mathbf{x} \in V_j\right\}.$$

The dimension of $\tilde{V}_j$ is $\tilde{d}_j$, and if $\mathbf{Y}$ is any $n \times n$ matrix, then there exist orthonormal vectors $\mathbf{p}_1, \ldots, \mathbf{p}_i$ in $\mathbb{R}^n$ such that $\mathbf{X}\mathbf{Y}\mathbf{p}_j \in V_j$, or $\mathbf{Y}\mathbf{p}_j \in \tilde{V}_j$ for $1 \leq j \leq i$. If the lemma were true for $n = k$, then it would imply that there would exist a j such that $\tilde{d}_j \geq i$. This would imply that $n + d_j - k \geq i$ or $d_j \geq k - n + i$, so it suffices to prove the lemma for $n = k$.

If $n = k$ and $d_j < i$ for $1 \leq j \leq i$, then (1) would imply that there exists a subspace U of $\mathbb{R}^n$ of dimension $n - i + 1$ such that $V_j \cap U = \{0\}$ for $1 \leq j \leq i$. If W is the subspace of dimension $i - 1$ that is perpendicular to U, then (3) would imply that for every $\delta > 0$, there exists an $\gamma_j > 0$ such that for all $\gamma_j > \varepsilon > 0$ if $\mathbf{X}_{W,\varepsilon}^{-1}\mathbf{v} \in V_j$ and $\|\mathbf{v}\| = 1$, then $\mathbf{v} \in S_{W,\delta}$. But then (2) would contradict the fact that there must exist orthonormal vectors $\mathbf{p}_1, \ldots, \mathbf{p}_i$ in $\mathbb{R}^n$ such that $\mathbf{X}_{W,\varepsilon}^{-1}\mathbf{p}_\mathbf{j} \in V_j$ for $1 \leq j \leq i$. So $d_j \geq i$ for some j as required by the lemma when $n = k$.

All that remains to be shown is (1), (2), and (3).

(1) Because $\dim(U + V_j) = \dim(U) + \dim(V_j) - \dim(V_j \cap U)$, (1) is equivalent to showing that there exists a subspace U of dimension d such that

$$\dim\left(U + V_j\right) = \min\left\{\dim\left(U\right) + \dim\left(V_j\right), k\right\}. \tag{20}$$

When U is of dimension 0, (20) is trivially true. Assume that there exists a subspace U of dimension $d-1$ for which (20) holds. We construct a subspace $\tilde{U}$ of dimension d for which (20) holds. For those j for which $\dim(U+V_j)<k$, the subspace $U+V_j$ is of measure zero in $\mathbb{R}^k$. So, there exists a $\mathbf{u}\notin U$ such that $\mathbf{u}\notin U+V_j$ for all j such that $\dim(U+V_j)<k$. Define $\tilde{U}$ to be the subspace spanned by U and $\mathbf{u}$. For those j for which $\dim(U+V_j)<k$, we have that $\dim(\tilde{U}+V_j)=\dim(U+V_j)+1$, and for those j for which $\dim(U+V_j)=k$, we have that $\dim(\tilde{U}+V_j)=k$. Since $\dim(\tilde{U})=\dim(U)+1$, verifying that (20) holds for $\tilde{U}$ is easy.

(2) Choose $\delta<1/(i\sqrt{n})$. Suppose there were $\mathbf{v_1},\ldots,\mathbf{v_i}$ in $S_{W,\delta}$ that were orthonormal. Since the $\mathbf{v}_j$ are in $S_{W,\delta}$, write $\mathbf{v_j}=\mathbf{w_j}+\delta\mathbf{u_j}$ where $\mathbf{w_j}\in W$, $\mathbf{u_j}$ is perpendicular to W, and $\|\mathbf{u_j}\|<1$. Let $\mathbf{X}$ be the $n\times i$ matrix $[w_1,\ldots,w_i]$, let $\mathbf{Y}$ be the $n\times i$ matrix $[\mathbf{v}_1,\ldots,\mathbf{v}_j]$, and let $\mathbf{Z}$ be the $n\times i$ matrix $[\mathbf{u}_1,\ldots,\mathbf{u}_i]$. Because the $\mathbf{w_j}$ are in W and the $\mathbf{u_j}$ are perpendicular to W, $\mathbf{X}'\mathbf{Z}=0$, and $\mathbf{Z}'\mathbf{X}=0$. Because the $\mathbf{v_j}$ are orthonormal, $\mathbf{Y}'\mathbf{Y}=\mathbf{I}_i$. So, $\mathbf{I}_i=\mathbf{X}'\mathbf{X}+\delta^2\mathbf{Z}'\mathbf{Z}$. Because the $\mathbf{w_j}$ are in a $i-1$ dimensional space, the matrix $\mathbf{X}'\mathbf{X}$ is singular, and there is a $\mathbf{v}\in\mathbb{R}^i$ of length one such that $\mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v}=0$. Because elements of $\mathbf{Z}$ and $\mathbf{v}$ are less than or equal to one in absolute value, each element of $\mathbf{Z}\mathbf{v}$ is less than or equal to i in absolute value. Thus $1=\mathbf{v}'\mathbf{v}=\delta^2\mathbf{z}'\mathbf{Z}'\mathbf{Z}v\leq\delta^2ni^2<1$, which is a contradiction. Therefore, $\mathbf{tv_1},\ldots,\mathbf{v_i}$ in $S_{W,\delta}$ that are orthonormal cannot exist.

(3) If Claim (3) were not true, then there would exist a $\delta>0$ and a sequence of $\mathbf{v}_\ell$ and ε_ℓ such that the ε_ℓ tend to zero and $\mathbf{X}^{-1}_{W,\varepsilon_\ell}\mathbf{v}_\ell\in V$, $\|\mathbf{v}_\ell\|=1$, and $\mathbf{v}_\ell\notin S_{W,\delta}$. We can write $\mathbf{v}_\ell=\mathbf{u}_\ell+\mathbf{w}_\ell$, where $\mathbf{u}_\ell\in U$ and $\mathbf{w}_\ell\in W$. Since $\mathbf{X}^{-1}_{W,\varepsilon_\ell}\mathbf{v}_\ell=\frac{1}{\varepsilon_\ell}\mathbf{u}_\ell+\mathbf{w}_\ell\in V$, we have that $\mathbf{u}_\ell+\varepsilon_\ell\mathbf{w}_\ell\in V$. Since $\|\mathbf{v}_\ell\|=1$, $\mathbf{u}_\ell$ and $\mathbf{w}_\ell$ are orthogonal, and $\mathbf{v}_\ell\notin S_{W,\delta}$, we have $\|\mathbf{w}_\ell\|\leq 1$ and $\delta\leq\|\mathbf{u}_\ell\|\leq 1$. Thus, $\lim_{\ell\to\infty}\varepsilon_\ell\mathbf{w}_\ell=0$, and hence some subsequence of the $\mathbf{u}_\ell$ converges to a non-zero element of $U\cap V$, which is a contradiction. ||

Lemma 9. *If the SVAR model is exactly identified, then $q_j=n-j$ for $1\leq j\leq n$.*

Proof. We proceed by contradiction. Suppose that it is not the case that $q_j=n-j$ for $1\leq j\leq n$. Since the model is exactly identified, Rothenberg's (1971) necessary condition for local identification implies that the total number of restrictions, $\sum_{i=1}^n q_i$, must be at least $(n-1)n/2$. Assuming that it is not the case that $q_j=n-j$ for $1\leq j\leq n$, there must be at least one index i with $q_i>n-i$. Since the q_j are in decreasing order, it must be the case that $q_j>n-i$ for $1\leq j\leq i$. If we define V_j to be the null space of the matrix $\mathbf{Q}_j$, then the dimension of V_j will be $n-q_j<i$. By Lemma 8, there exists a $k\times n$ matrix $\mathbf{X}$ for which there are no orthonormal vectors $\mathbf{p}_1,\ldots,\mathbf{p}_i$ in $\mathbb{R}^n$ such that $\mathbf{X}\mathbf{p}_j\in V_j$ for $1\leq j\leq i$. Let $M(k,n)$ be the set of all $k\times n$ matrices and let

$$\tilde{H}=\{\mathbf{X}\in M(k,n)\mid \mathbf{XP}\notin\tilde{\mathcal{R}}\text{ for all }\mathbf{P}\in O(n)\}. \tag{21}$$

We have shown that $\tilde{H}$ is non-empty. The proof will be complete if we can show that $\tilde{H}$ is open. To see this, note that since $f(U)$ is dense, $\tilde{H}\cap f(U)$ would be a non-empty open set, as would $f^{-1}(\tilde{H})$. Since open sets are of positive measure, the fact that the SVAR is exactly identified is contradicted.

To show that $\tilde{H}$ is open, it suffices to show that if the sequence $\mathbf{X}_j\notin\tilde{H}$ converges to $\mathbf{X}$, then $\mathbf{X}\notin\tilde{H}$. If $\mathbf{X}_j\notin\tilde{H}$ then there would exist $\mathbf{P}_j\in O(n)$ such that $\mathbf{X}_j\mathbf{P}_j\in\tilde{\mathcal{R}}$. Since $O(n)$ is a compact set, some subsequence of the $\mathbf{P}_j$ converges to some orthogonal matrix $\mathbf{P}$. Since $\tilde{\mathcal{R}}$ is a closed subset, $\mathbf{XP}\in\tilde{\mathcal{R}}$ as desired. ||

Proof of Theorem 7. This result follows directly from Lemmas 7 and 9. ||

Proof of Theorem 6. If the SVAR is exactly identified, then by Theorem 7, $q_j=n-j$ for $1\leq j\leq n$. Clearly, the number of restrictions, $\sum_{i=1}^n q_i$, is equal to $(n-1)n/2$ and by Lemma 6, the rank condition in Theorem 1 is satisfied for some $(\mathbf{A}_0,\mathbf{A}_+)\in R$.

On the other hand, since $\text{rank}(\mathbf{M}_j(\mathbf{X}))\leq q_j+j$, if rank condition in Theorem 1 is satisfied for some $(\mathbf{A}_0,\mathbf{A}_+)\in R$, it must be the case that $q_j\geq n-j$ for $1\leq j\leq n$. If it is also the case that the total number of restrictions is equal to $(n-1)n/2$, then $q_j=n-j$ for $1\leq j\leq n$. So, by Theorem 7, the SVAR is exactly identified. ||

APPENDIX C. PROOF OF THEOREM 4

Proof of Theorem 4. The first step in the proof is an application of Algorithm 1 to a certain structural parameter point $(\mathbf{A}_0,\mathbf{A}_+)$. We show that if

$$V_i=\left\{\mathbf{v}\in\mathbb{R}^n\mid\mathbf{Q}_i\mathbf{v}=0\right\},$$

then $V_1\subset\ldots\subset V_n$. If this were not the case, then let j be the first index such that $V_{j-1}\not\subseteq V_j$. We could then find linearly independent $\mathbf{a_1},\ldots,\mathbf{a_n}$ such that $\mathbf{a}_i\in V_i$ for $1\leq i<j$ and $\mathbf{a}_j,\mathbf{a}_{j+1}\in V_j$. Let $(\mathbf{A}_0,\mathbf{A}_+)$ be any structural

parameter point with $\mathbf{A}_0 = [\mathbf{a_1} \ \ldots \ \mathbf{a_n}]$. Applying Algorithm 1 will produce an orthogonal $\mathbf{P}$ such that $(\mathbf{A}_0\mathbf{P}, \mathbf{A}_+\mathbf{P}) \in R$ and, by the hypotheses of the theorem, this $\mathbf{P}$ must be unique. But $\mathbf{P}$ will be unique if and only if the matrix $\tilde{\mathbf{Q}}_i$ defined by (10) is of rank $n-1$ for $1 \leq i \leq n$. Because $\mathbf{a_i} \in V_i$ for $1 \leq i < j$, it must be the case that for $1 \leq i < j$, $\mathbf{Pe_i} = \pm\mathbf{e_i}$, where $\mathbf{e_i}$ is the i-th column of $\mathbf{I}_n$. But $\tilde{\mathbf{Q}}_j$ cannot be of rank $n-1$ because both $\tilde{\mathbf{Q}}_j\mathbf{e}_j$ and $\tilde{\mathbf{Q}}_j\mathbf{e}_{j+1}$ will be zero. This result is a contradiction and so it must be the case that $V_1 \subset \ldots \subset V_n$.

To complete the proof, we show that if $V_1 \subset \ldots \subset V_n$, then the system is triangular. Because $V_1 \subset \ldots \subset V_n$, there exist linearly independent $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that $\mathbf{v}_1, \ldots, \mathbf{v}_i$ forms a basis for V_i. If we define $\mathbf{P}_1 = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathbf{P}_0$ to the matrix with ones along the antidiagonal, then $(\mathbf{A}_0, \mathbf{A}_+) \in R$ if and only if $\mathbf{P}_1 f(\mathbf{A}_0, \mathbf{A}_+)\mathbf{P}_0 = \mathbf{P}_1\mathbf{A}_0\mathbf{P}_0$ is lower triangular. Thus, the system is triangular by Definition 6.3. ||

Acknowledgements. We are grateful to the referees and the editor (Bernard Salanié), whose detailed comments have led to a significant improvement in this paper. We wish to thank Jushan Bai, Fabio Canova, John Chao, Xiaohong Chen, Tim Cogley, Wouter Den Haan, Jean-Marie Dufour, Jon Faust, Cheng Hsiao, John Geweke, Roger Moon, Jim Nason, Chris Otrok, Adrian Pagan, Ingmar Prucha, John Rust, and Harald Uhlig for helpful discussions and comments. The current version of this paper draws heavily from the two unpublished manuscripts "Markov-Switching Structural Vector Autoregressions: Theory and Application" by Rubio-Ramírez, Waggoner, and Zha and "Identification Issues in Vector Autoregressions" by Waggoner and Zha. The views expressed are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

REFERENCES

BEKKER, P. A. and POLLOCK, D. (1986), "Identification of Linear Stochastic Models with Covariance Restrictions", *Journal of Econometrics,* **31**, 179–208.

BENATI, L. and MUMTAZ, H. (2006), "U.S. Evolving Macroeconomic Dynamics: A Structural Investigation" (Unpublished Manuscript, European Central Bank and Bank of England).

BERNANKE, B. S. and MIHOV, I. (1998), "Measuring Monetary Policy", *Quarterly Journal of Economics,* **113** (3), 869–902.

BLANCHARD, O. J. and QUAH, D. (1993), "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review,* **83** (3), 655–673.

CANOVA, F. and DENICOLÓ, G. (2002), "Monetary Disturbances Matter for Business Fluctuations in the G-7", *Journal of Monetary Economics,* **49** (6), 1121–1159.

CANOVA, F. and GAMBETTI, L. (2004), "Structural Changes in the US Economy: Bad Luck or Bad Policy?" (Manuscript, Universitat Pompeu Fabra, August 2004).

CHRISTIANO, L., EICHENBAUM, M. and EVANS, C. (2005), "Nominal Rigidities and the Dynamics Effects of a Shock to Monetary Policy", *Journal of Political Economy,* **113**, 1–45.

CHRISTIANO, L. J., EICHENBAUM, M. and EVANS, C. L. (1996), "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds", *Review of Economics and Statistics,* **78** (1), 16–34.

COCHRANE, J. H. (2006), "Identification and Price Determination with Taylor Rules: A Critical Review" (Manuscript, University of Chicago, June 2006).

COGLEY, T. and SARGENT, T. (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.", *Review of Economic Dynamics,* **8**, 262–302.

DHRYMES, P. J. (1978), *Introductory Econometrics* (New York: Springer-Verlag).

DHRYMES, P. J. (1994), *Topics in Advanced Econometrics* (New York: Springer-Verlag).

EICHENBAUM, M. and EVANS, C. (1995), "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates", *Quarterly Journal of Economics,* **110**, 975–1009.

FAUST, J. (1998), "The Robustness of Identified Var Conclusions About Money", *Carnegie-Rochester Conference Series on Public Policy,* **48**, 207–244.

FERNÁNDEZ-VILLAVERDE, J. and RUBIO-RAMÍREZ, J. F. (2004), "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach", *Journal of Econometrics,* **123**, 153–187.

FISHER, F. M. (1966), *The Identification Problem in Econometrics* (New York: McGraw-Hill).

FRY, R. and PAGAN, A. (2007), "Some Issues in Using Sign Restrictions for Identifying Structural Vars" (National Centre for Econometric Research Working Paper 14, April 2007).

FUBAC, M., WAGGONER, D. F. and ZHA, T. (2007), "Local and Global Identification of DSGE Models: A Simultaneous-Equation Qpproach" (Manuscript, Federal Reserve Bank of Atlanta).

GALÍ, J. (1992), "How Well Does the IS-LM Model Fit Postwar U.S. Data?" *Quarterly Journal of Economics,* **107** (2), 709–738.

GALÍ, J. (1999), "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?", *American Economic Review,* **89** (1), 249–271.

GAMBETTI, L. and CANOVA, F. (2008), "Structural Dynamics of US Output and Inflation: What Explains the Changes?" *Journal of Money, Credit, and Banking,* **40**, 369–388.

GEWEKE, J. (1986), "Exact Inference in the Inequality Constrained Normal Linear Regression Model", *Journal of Applied Econometrics,* **1**, 127–142.
GEWEKE, J. (1999), "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication", *Econometric Reviews,* **18**, 1–126.
GEWEKE, J. (2005), *Contemporary Bayesian Econometrics and Statistics* (Hoboken, NJ: John Wiley & Sons).
GIANNINI, C. (1992), *Topics in Structural VAR Econometrics* (New York: Springer-Verlag).
GORDON, D. B. and LEEPER, E. M. (1994), "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification", *Journal of Political Economy,* **102** (6), 1228–1247.
GREENE, W. H. (1993), *Econometric Analysis*, 2nd edn (New York: Macmillan).
HAMILTON, J. D. (1994), *Times Series Analysis* (Princeton: Princeton University Press).
HAMILTON, J. D., WAGGONER, D. F. and ZHA, T. (2007), "Normalization in Econometrics", *Econometric Reviews,* **26** (2-4), 221–252.
HAUSMAN, J. A. and TAYLOR, W. E. (1983), "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation", *Econometrica,* **51** (5), 1527–1549.
HSIAO C., (1983), "Identification", in Z. Griliches and M. M. Intrilligator, (eds.) *Handbook of Econometrics I*, Volume 1 (North-Holland) 224–283.
INOUE, A. and KILIAN, L. (2002), "Bootstrapping Autoregressive Processes with Possible Unit Roots", *Econometrica,* **70** (1), 377–391.
ISKREV, N. (2007), "How Much Do We Learn From the Estimation of DSGE Models? A Case Study of Identication Issues in a New Keynesian Business Cycle Model" (Manuscript, University of Michigan, October 2007).
KILIAN, L. (1998), "Small-Sample Confidence Intervals for Impulse Response Functions", *Review of Economics and Statistics,* **80**, 186–201.
KING, R. G., PLOSSER, C. I., STOCK, J. H. and WATSON, M. W. (1991), "Stochastic Trends and Economic Fluctuations", *American Economic Review,* **81** (4), 819–840.
LEEPER, E. M., SIMS, C. A. and ZHA, T. (1996), "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity,* **2**, 1–78.
LEEPER, E. M. and ZHA, T. (2003), "Modest Policy Interventions", *Journal of Monetary Economics,* **50** (8), 1673–1700.
PEERSMAN, G. and SMETS, F. (2003), *The Monetary Transmission Mechanism in the Euro Area: More Evidence from VAR Analysis* (Cambridge: Cambridge University Press).
PESAVENTO, E. and ROSSI, B. (2006), "Small-Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons", *Journal of Applied Econometrics,* **21** (8), 1135–1155.
PRIMICERI, G. E. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy", *Review of Economic Studies,* **72**, 821–852.
ROTHENBERG, T. J. (1971), "Identification in Parametric Models", *Econometrica,* **39** (3), 577–591.
RUBIO-RAMÍREZ, J. F., WAGGONER, D. F. and ZHA, T. (2005), "Markov-Switching Structural Vector Autoregressions: Theory and Application" (Federal Reserve Bank of Atlanta, Working Paper 2005-27).
RUBIO-RAMÍREZ, J. F., WAGGONER, D. F. and ZHA, T. (2008), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference" (Federal Reserve Bank of Atlanta Working Paper 2008-18, September 2008).
SARGENT, T. J. (1976), "The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics", *Journal of Political Economy,* **84** (3), 631–640.
SIMS, C. A. (1980), "Macroeconomics and Reality", *Econometrica,* **48** (1), 1–48.
SIMS, C. A. (1986), "Are Forecasting Models Usable for Policy Analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review,* **10**, 2–16.
SIMS, C. A. (2005), "Dummy Observation Priors Revisited" (Manuscript, Princeton University).
SIMS, C. A. and ZHA, T. (1998), "Bayesian Methods for Dynamic Multivariate Models", *International Economic Review,* **39** (4), 949–968.
SIMS, C. A. and ZHA, T. (1999), "Error Bands for Impulse Responses", *Econometrica,* **67** (5), 1113–1155.
SIMS, C. A. and ZHA, T. (2006*a*), "Does Monetary Policy Generate Recessions?" *Macroeconomic Dynamics,* **10**, 231–272.
SIMS, C. A. and ZHA, T. (2006*b*), "Were There Regime Switches in US Monetary Policy?", *American Economic Review,* **96**, 54–81.
SMETS, F. and WOUTERS, R. (2003), "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area", *Journal of European Economic Association,* **1** (5), 1123–1175.
SPIVAK, M. (1965), *Calculus on Manifolds: A Modern Approach to Classical Thoerems of Advanced Calculus* (New York: Benjamin/Cummings).
STEWART, G. W. (1980), "The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators", *SIAM Journal on Numerical Analysis,* **17** (3), 403–409.
UHLIG, H. (1997), "Bayesian Vector Autoregressions with Stochastic Volatility", *Econometrica,* **65**, 59–73.
UHLIG, H. (2005), "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure", *Journal of Monetary Economics,* **52** (2), 381–419.
WAGGONER, D. F. and ZHA, T. (2003), "Likelihood Preserving Normalization in Multiple Equation Models", *Journal of Econometrics,* **114** (2), 329–347.
ZHA, T. (1999), "Block Recursion and Structural Vector Autoregressions", *Journal of Econometrics,* **90** (2), 291–316.