A New Method for Working with Sign Restrictions in SVARs

S Ouliaris
A R Pagan

Working Paper #105
May 2015
A New Method for Working With Sign Restrictions in SVARs

S. Ouliaris* A.R. Pagan †

April 11, 2015

Contents

1 Introduction 2

2 Two Simple Structural Models and Their Sign Restrictions 4

3 Two Methods for Using Sign Restriction Information 5
   3.1 The SRR Method .......................... 5
   3.2 The SRC method .......................... 7

4 The SRC and SRR Methods Applied to the Market Model 8
   4.1 The SRR Method .......................... 8
   4.2 The SRC Method .......................... 9
   4.3 Comparing the SRR and SRC Methodologies ........... 10

5 Comparing SRC and SRR With Some Simulated and Actual Data 11
   5.1 A Simulated Market Model .................. 11
   5.2 The Small Macro Model with Transitory Shocks ........ 11

6 Combining Parametric and Sign Restrictions 13

*Institute for Capacity Development, International Monetary Fund
†Melbourne Institute of Applied Economic and Social Research and School of Economics, University of Sydney, adrian.pagan@sydney.edu.au
Abstract

Structural VARs are used to compute impulse responses to shocks. One problem that has arisen involves the information needed to perform this task i.e. how are the shocks separated into those representing technology, monetary effects etc. Increasingly the signs of impulse responses are used for this task. However it is often desirable to impose some parametric assumption as well e.g. that monetary shocks have no long-run impact on output. Existing methods for combining sign and parametric restrictions are not well developed. In this paper we provide a relatively simple way to allow for these combinations and show how it works in a number of different contexts.

1 Introduction

Structural Vector Autoregressions (SVARs) have become a standard way of modelling macroeconomic series. A SVAR of order p in n variables $y_t$ is

$$A_0 y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + \Omega \varepsilon_t,$$

where $\Omega$ is $diag\{\sigma_1, ..., \sigma_n\}$ and $A_0$ has unity on its diagonal. Associated with the SVAR is the Moving Average representation

$$y_t = C(L)\varepsilon_t = C_0 \varepsilon_t + C_1 \varepsilon_{t-1} + ...,$$

where the $C_j$ are the impulse responses of $y_{t+j}$ to a unit shock in $\varepsilon_t$. When some of the variables in $y_t$ are differences of $I(1)$ variables, $z_t$, the long-run impulse responses in $C(1)$ show the response of $z_\infty$ to variations in $\varepsilon_t$.

Because the SVAR is a set of structural equations there is a limit to the number of parameters in $A_0$ that can be estimated. A variety of methods
have emerged to deal with this problem. Short-run restrictions set some of the elements of $A_0$ to zero. When some of the members of $y_t$ are differences in $I(1)$ variables long-run restrictions may be imposed, and these translate into restrictions on $C(1)$ i.e. constraints between the $A_j$. It is also possible to envisage restrictions upon the impulse responses $C_j$ themselves. All of these methods are examples of how parametric restrictions can be used to distinguish between the shocks in $\varepsilon_t$.

Many extensions have been made to the SVAR modelling framework described above. One is to allow $p$ to be infinite - see Lütkepohl and Saikkonen (1997). Another, motivated by the fact that estimation of $A_0$ requires instruments, finds these by using information about structural change in the parameters of the SVAR - see Herwartz and Lütkepohl (2011) and Lanne and Lütkepohl (2008). A third has been the suggestion that the shocks $\varepsilon_t$ can be differentiated by the signs of the impulse responses in $C_j$, rather than by whether they have precise numerical values - see Faust (1998), Canova and De Nicoló (2002), Uhlig (2005) and Peersman (2005).

A method has evolved for utilizing sign restriction information that involves generating many sets of impulse responses and retaining those that satisfy the sign restrictions. To generate many impulse responses one starts with an initial or base set of shocks which are also uncorrelated, and then recombines them in such a way that the new shocks are uncorrelated while having a different set of impulse responses. Because it involves recombination we will refer to this as the SRR approach, where the SR indicates sign restrictions and the R is recombination.

In this paper a new approach is suggested. It recognizes that, once $n(n - 1)/2$ elements of $A_0$ are prescribed, the remaining coefficients can be estimated and impulse responses found. Hence, by varying the prescribed elements in a random way, it is possible to get a large set of impulse responses that correspond to the recombined ones in SRR. The method is called SRC, where the C comes from the fact that coefficients in $A_0$ are set. Once many impulse responses are found one can proceed just as for SRR and check if they satisfy the sign restrictions. Section 3 sets out the details of SRR and SRC more precisely.

In section 4 the methodologies of SRR and SRC are compared by using two simple models set out in section 2 - a market model and a small macro model. Then section 5 provides a numerical comparison using both simulated and actual data. Because the SRC method involves estimation of a system of structural equations it is relatively easy to use when faced with the need to
estimate SVARs that feature a mixture of sign and parametric restrictions. This is demonstrated with an application of the small macro model in section 6. Another use of the SRC method is to find estimates of the standard deviations of the shocks. SRR does not do this. However, by studying why it can be done with SRC, it is possible to find a procedure for obtaining the standard deviations in the context of SRR, and this is done in section 7. Finally, standard errors for the impulse responses for both SRR and SRC are discussed in section 8. Section 9 concludes.

2 Two Simple Structural Models and Their Sign Restrictions

We will use two simple models to illustrate the arguments. These are taken from Fry and Pagan (2011). One is a market model, which has the form

\[ q_t = \alpha p_t + \sigma_S \epsilon_{1t}, \]
\[ q_t = -\beta p_t + \sigma_D \epsilon_{2t}, \]

where \( q_t \) is quantity, \( p_t \) is price, and the shocks \( \epsilon_{jt} \) are \( n.i.d.(0,1) \) and uncorrelated with one another. The first curve might be a supply curve and the second a demand curve (implying that both \( \alpha \) and \( \beta \) are positive). Because lags are omitted from (1) and (2) this is a structural system, but not a SVAR. Nevertheless, it is useful to abstract from lags and this can be done without loss of generality.

We would expect that the signs of the contemporaneous responses of prices and quantity to positive demand and cost shocks would be those of Table 1.

<p>| Table 1: Sign Restrictions for the Market Model (Demand/Supply (Cost) Shocks) |</p>
<table>
<thead>
<tr>
<th>variable\shock</th>
<th>Demand</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_t )</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>( q_t )</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

A second one is a small macro model which involves an output gap \( (y_t) \), inflation \( (\pi_t) \), and a policy interest rate \( (i_t) \), with its SVAR form being
\[ y_t = \xi_{t-1}'y + \beta_{yi}i_t + \beta_{y\pi}\pi_t + \varepsilon_{yt} \]
\[ \pi_t = \xi_{t-1}'\pi + \beta_{\pi i}i_t + \beta_{\pi y}y_t + \varepsilon_{\pi t} \]
\[ i_t = \xi_{t-1}'i + \beta_{iy}y_t + \beta_{i\pi}\pi_t + \varepsilon_{it}, \]

where \( \xi'_{t-1} = (y_t, \pi_t, i_t) \).

We might expect the sign restrictions on the contemporaneous responses for positive shocks in this model to be those in Table 2.

<table>
<thead>
<tr>
<th>Table 2: Sign Restrictions for Macro Model Shocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable\shock</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>( y_t )</td>
</tr>
<tr>
<td>( \pi_t )</td>
</tr>
<tr>
<td>( i_t )</td>
</tr>
</tbody>
</table>

3 Two Methods for Using Sign Restriction Information

In the introduction we mentioned two methods for using sign restrictions in SVARs, SRR and SRC, and we noe set these out in more detail.

3.1 The SRR Method

The key to the SSR method is the selection of a set of base shocks \( \eta_t \) that are uncorrelated and which have zero mean and unit variance. One way of getting these is to use the estimated structural shocks from assuming that the system is recursive (this may be totally wrong but all we are trying to do is get a set of base shocks that are uncorrelated). In that case

\[ A_{0}^{\text{rec}} z_{t} = A_{1} z_{t-1} + \varepsilon_{t}^{R}, \]

where \( A_{0}^{\text{rec}} \) is a triangular matrix with unity on the diagonals (the equations are normalized) and the \( \varepsilon_{t}^{R} \) are the recursive system structural shocks. Then, the estimated standard deviations of \( \varepsilon_{t}^{R} \) can be used to produce \( \tilde{\varepsilon}_{j}^{R} = \frac{\varepsilon_{j}^{R}}{\text{std}(\varepsilon_{j}^{R})} \).
and the $\tilde{\epsilon}_R^j$ will have unit variances. Consequently, if $\eta_t$ is set equal to $\tilde{\epsilon}_t^R$, it can be thought of as i.i.d. $(0, I_n)^1$

Once $\eta_t$ is found there is an MA structure that determines the impulse responses. Thus, for the recursive model,

$$z_t = C^{\text{recur}}(L)\tilde{\epsilon}_t^R$$

$$= C^R(L)\tilde{\epsilon}_t^R = C^R(L)\eta_t,$$

showing that the impulse responses to the shocks $\eta_t$ are different to the original set. Given this feature the methodology of SRR involves continually forming new shocks ($\eta^*_t$) by combining those from the base shocks ($\eta_t$) in such a way that the new shocks remain uncorrelated i.e. $\eta^*_t = Q\eta_t$, where the $n \times n$ matrix $Q$ is required to have the property

$$Q'Q = I_n, QQ' = I_n. \tag{3}$$

It is crucial to observe that the new shocks $\eta^*_t$ need not come from a recursive system even if $\eta_t$ does.

There are a number of ways to find a $Q$ with the requisite properties - see Fry and Pagan (2011). A useful method for choosing $Q$ which suits the exposition of this paper is a Givens rotation. When there are two variables ($n = 2$) this has the structure

$$Q = \begin{bmatrix} \cos \lambda & -\sin \lambda \\ \sin \lambda & \cos \lambda \end{bmatrix}, 0 \leq \lambda \leq \pi$$

Consequently, many different $Q$ matrices and impulse responses can be generated by using a range of values for $\lambda$. Since $\lambda$ lies between 0 and $\pi$ we could just set up a grid of values. An alternative is to use a random number generator, drawing $\lambda$ (say) from a uniform density over 0 to $\pi$. Let the $m$'th draw produce $\lambda^{(m)}, m = 1, ..., M$. Once a $\lambda^{(m)}$ is available then $Q^{(m)}$ can be computed, and there will be $M$ models with impulse response functions $C^{(m)}_j$.

Of course, although all these models are distinguished by different numerical values for $\lambda$, they are are observationally equivalent, in that they produce an exact fit to the variance of the data on $z_t$.\(^2\) Only those impulse responses (and associated $Q^{(m)}$) producing shocks that agree with the maintained sign restrictions would be retained.

---

\(^1\)This is not the only way of getting $\eta_t$. Fry and Pagan (2010) discuss others.

\(^2\)This statement assumes a zero mean for $z_t$. 

6
In the context of a 3 variable VAR (as in the small macro model) a $3 \times 3$ Givens matrix $Q_{12}$ has the form

$$Q_{12} = \begin{bmatrix} \cos \lambda & -\sin \lambda & 0 \\ \sin \lambda & \cos \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

i.e. the matrix is the identity matrix in which the block consisting of the first and second columns and rows has been replaced by cosine and sine terms and $\lambda$ lies between 0 and $\pi$. There are three possible Givens rotations for a three variable system - the others being $Q_{13}$ and $Q_{23}$. Each of the $Q_{ij}$ depends on a separate parameter $\lambda_k$ ($k = 1, \ldots, 3$). In practice most users of the approach have adopted the multiple of the basic set of Givens matrices as $Q$ e.g. in the three variable case we would use

$$Q_G(\lambda) = Q_{12}(\lambda_1) \times Q_{13}(\lambda_2) \times Q_{23}(\lambda_3).$$

It’s clear that $Q_G$ is orthogonal and so shocks formed as $\eta_t^* = Q_G\eta_t$ will be uncorrelated. Because the matrix $Q_G$ above depends upon three different $\lambda_k$ one could draw each $\lambda_k$ from a $U(0, \pi)$ density function.

As just intimated the process doesn’t end with just one $Q$. It is iterated to produce many impulse responses by varying $Q$. Each time they are tested for whether they obey the postulated sign restrictions. Thus, this leads to the following *modus operandi* for SRR.

1. Start with a set of uncorrelated shocks $\eta_t$ that have $I_n$ as their covariance matrix.
2. Generate a new set of shocks $\eta_t^* = Q\eta_t$ using a $Q$ with the properties $Q'Q = QQ' = I_n$.
3. Compute the IRF’s for this set of shocks.
4. If they have the correct signs RETAIN them. If not draw another $Q$.

### 3.2 The SRC method

Because the shocks are connected to a SVAR $\frac{n(n-1)}{2}$ elements of $A_0$ will be estimated using the restriction that the shocks are uncorrelated and $A_0$ has
unity on the diagonal. Therefore there will be \( \frac{n(n-1)}{2} \) non-estimable parameters in \( A_0 \). These need to be fixed to some values if estimation is to proceed. The idea behind the SRC approach is to choose some values for the non-estimable parameters in \( A_0 \) and to then estimate the remainder with a method which ensures that the shocks are uncorrelated. Because there is no unique way to set values for the non-estimable parameters, there will need to be many values generated for them and these will produce different \( A_0 \) and hence impulse responses i.e. it performs the same task as varying \( Q \) values in SRR. So the key to the methodology resides in generating many values for the non-estimable parameters, and these will be taken to depend upon some quantities designated as \( \theta \). Broadly we will find values for the non-estimable parameters by generating candidate values for \( \theta \) from a random number generator. The context may determine exactly how that would be done. Once again the models found with different values of \( \theta \) are observationally equivalent since the SVAR is exactly identified.

4 The SRC and SRR Methods Applied to the Market Model

The market model of section 2 can be represented as

\[
\begin{align*}
aq_t &= bp_t + \varepsilon_{1t} \\
cq_t &= dp_t + \varepsilon_{2t}.
\end{align*}
\]

The corresponding impulse responses to these shocks will be \( \begin{bmatrix} a & -b \\ c & -d \end{bmatrix}^{-1} \).

4.1 The SRR Method

In SRR one way to initiate the process is to start with a recursive model. For the market model this could be

\[
\begin{align*}
q_t &= s_1 \eta_{1t} \\
p_t &= \phi q_t + s_2 \eta_{2t}.
\end{align*}
\]

Data is available on \( q_t \) and \( p_t \) and the \( \eta_{jt} \) are \( n.i.d(0, 1) \), with \( s_j \) being the standard deviations of the errors for the two equations. The first stage of
SRR is then implemented by applying some weighting matrix $Q$ to the initial shocks $\eta_{1t}$ and $\eta_{2t}$, so as to produce new shocks $\eta^*_{1t}$ and $\eta^*_{2t}$, i.e. $\eta^*_t = Q\eta_t$.

Using the Givens matrix $Q = \begin{bmatrix} \cos \lambda & -\sin \lambda \\ \sin \lambda & \cos \lambda \end{bmatrix}$ the new shocks $\eta^*_t = Q\eta_t$ will be

$$\begin{align*}
\cos \lambda \eta_{1t} - \sin \lambda \eta_{2t} &= \eta^*_{1t} \\
\sin \lambda \eta_{1t} + \cos \lambda \eta_{2t} &= \eta^*_{2t}.
\end{align*}$$

and this can be written in the same form as (4)-(5) by setting

$$\begin{align*}
a &= \cos \lambda/s_1 + (\sin \lambda)(\phi/s_2), \\
b &= \sin \lambda/s_2, \\
c &= \sin \lambda/s_1 - (\cos \lambda)(\phi/s_2), \\
d &= -\cos \lambda/s_2
\end{align*}$$

(8)

$$\varepsilon_{jt} = \eta^*_{jt}$$

(10)

This provides an alternative view of what the SRR method does, namely it generates many impulse responses by expressing the $A_0$ coefficients of the SVAR model in terms of $\lambda$, and then varying $\lambda$ over the region $(0, \pi)$. Once the impulse responses are found sign restrictions are applied to determine which are to be retained. So we are generating many impulse responses by making the market model parameters $A_0$ depend upon $\lambda$ and the data (through $\phi$, $s_1$ and $s_2$).

### 4.2 The SRC Method

Rather than expressing the model parameters in terms of $\lambda$, consider the possibility of going back to (1) and making the coefficient $\alpha$ (the non-estimable one) a function of $\theta$ as $\alpha = \frac{\theta}{(1-\text{abs}(\theta))}$, where $\theta$ is drawn from a uniform density over $(-1,1)$. Once a value of $\theta$ is found this will fix $\alpha$. The estimable coefficients then need to be found from the data in such a way as to produce uncorrelated shocks.

After setting $\theta$ to some generated value $\theta^*$ SRC proceeds in the following way

1. Form residuals $\hat{\varepsilon}^*_{1t} = q_t - \alpha(\theta^*)p_t$.

2. Estimate $\sigma_1$ with $\hat{\sigma}_1^*$, the standard deviation of these residuals.
3. Using $\hat{\varepsilon}_{1t}^*$ as an instrument for $p_t$ estimate $\beta$ by Instrumental Variables (IV) to get $\hat{\beta}^*$.

4. Using $\hat{\beta}^*$ form the residuals $\hat{\varepsilon}_{2t}^* = q_t + \hat{\beta}^* p_t$. The standard deviation of these, $\hat{\sigma}_{2t}^*$, will estimate the standard deviation of the second shock. By the nature of the estimation procedure the shocks $\hat{\varepsilon}_{1t}^*$ and $\hat{\varepsilon}_{2t}^*$ are orthogonal.

Using earlier results, the contemporaneous impulse responses to one standard deviation shocks will be 

$$
\begin{bmatrix}
1 & -\alpha(\theta^*) \\
1 & \hat{\beta}^*
\end{bmatrix}^{-1}
\begin{bmatrix}
\hat{\sigma}_1^* & 0 \\
0 & \hat{\sigma}_2^*
\end{bmatrix}.
$$

Accordingly, just as happened with $\lambda$ in the SRR approach, we can vary $\theta$ and thereby generate many impulse responses. These are directly comparable with the impulse responses generated by SRR, except that they all depend upon $\theta$ and the data (via the IV estimation) rather than $\lambda$ and the data.

### 4.3 Comparing the SRR and SRC Methodologies

It is worth looking closer at these two methods in terms of the market model above. A number of points emerge.

(i) $\theta$ will normally be chosen so as to get a range of variation in $\alpha$ that is $(-\infty, \infty)$. As mentioned above this can be done by drawing $\theta$ from a uniform (-1,1) density and then setting $\alpha = \frac{\theta}{1 - \text{abs}(\theta)}$. By comparison $\lambda$ is drawn from a uniform density over $(0, \pi)$, because of the presence of $\lambda$ in the harmonic terms. In both approaches one has to decide upon the number of trial values of $\theta$ and $\lambda$ to use i.e. how many sets of impulse responses are to be computed. We note that there may be cases where it is possible to bound the values of the non-estimable parameters and this would then have implications for how $\theta$ is generated (or possibly one would simply discard models for which the non-estimable parameters lay outside the bounds).

(ii) In a SVAR with $n$ variables and no parametric restrictions the number of $\lambda_j$ to be generated in the SRR method equals $n(n-1)/2$. Thus, for $n = 3$, three $\lambda_j$'s are needed. This is also true of the number of $\theta_j$ used in SRC. So problems arising from the dimensions of the system are the same for both methods. It should be noted however that, when parametric restrictions are also applied along with sign restrictions, the number of $\theta_j$ may be much smaller and this will be shown later. Such an outcome should be apparent because parametric restrictions increase the number of estimable parameters.
in \( A_0 \) and, since \( \theta_j \) relates to the non-estimable parameters, a smaller number of \( \theta_j \) need to be generated.

5 Comparing SRC and SRR With Some Simulated and Actual Data

5.1 A Simulated Market Model

To look more closely at these two methods we simulate data from the following market model

\[
q_t = 3p_t + \sqrt{2}\varepsilon_{2t} \\
q_t = -p_t + \varepsilon_{1t}.
\]

The true impulse responses for price and quantity (with the demand shock first and supply second) are \[
\begin{bmatrix}
.25 \\
.75
\end{bmatrix}
\begin{bmatrix}
-.3536 \\
.3536
\end{bmatrix}.
\]

Five hundred values for \( \theta \) and \( \lambda \) were generated from a uniform random number generator (over \((0, \pi)\) for \( \lambda \) and \((-1,1)\) for \( \theta \)) and the impulse responses were compared to the sign restrictions in Table 2. SRR generates impulses that are compatible with the sign restrictions 87.8% of the time and for SRC it is 85.4%. This is a high percentage but, since the model is correct, that is what would be expected. Inspecting these 500 impulse responses we find that the closest fit for each method to the true values was\(^3\)

\[
SRC = \begin{bmatrix}
.2484 \\
.7369
\end{bmatrix}, SRR = \begin{bmatrix}
.2472 \\
.7648
\end{bmatrix}.
\]

It is clear that among the 500 sets of responses for each method there is at least one that gives a good match to the true impulse responses. Changing the parameter values for the market model did not change this conclusion.

5.2 The Small Macro Model with Transitory Shocks

We will now look at the two methods in the context of the three variable small macro model. The variables in the system consist of three variables

\(^3\)We just use a simple Euclidean norm to define the closest match to the true values. The impulse responses are to a one standard deviation shock.
$y_{1t}, y_{2t}$ and $y_{3t}$, where $y_{1t}$ is an output gap, $y_{2t}$ is quarterly inflation, and $y_{3t}$ is a nominal interest rate. All variables are assumed to be $I(0)$ and that there are three transitory shocks - labelled productivity, demand and an interest rate. The expected signs of the contemporaneous impulse responses are given in Table 2.

The model fitted is the SVAR(1) with data from Cho and Moreno (2006).

\[ y_{1t} = a_{12}^0 y_{2t} + a_{13}^0 y_{3t} + a_{12}^1 y_{2t-1} + a_{13}^1 y_{3t-1} + a_{11}^1 y_{1t-1} + \varepsilon_{1t} \] (12)

\[ y_{2t} = a_{21}^0 y_{1t} + a_{23}^0 y_{3t} + a_{22}^1 y_{2t-1} + a_{23}^1 y_{3t-1} + a_{21}^1 y_{1t-1} + \varepsilon_{2t} \] (13)

\[ y_{3t} = a_{31}^0 y_{1t} + a_{32}^0 y_{2t} + a_{32}^1 y_{2t-1} + a_{33}^1 y_{3t-1} + a_{31}^1 y_{1t-1} + \varepsilon_{2t}. \] (14)

The SRR method begins by setting $a_{12}^0 = 0, a_{13}^0 = 0$ and $a_{23}^0 = 0$ to produce a recursive model, and then recombines the impulse responses found from this model using the $Q_G$ matrix that depends upon $\lambda_1, \lambda_2$ and $\lambda_3$. In contrast, the SRC method proceeds by first fixing $a_{12}^0$ and $a_{13}^0$ to some values and then computing residuals $\hat{\varepsilon}_{1t}$. After this (13) is estimated by fixing $a_{23}^0$ to some value and using $\hat{\varepsilon}_{1t}$ as an instrument for $y_{1t}$. Finally, the residuals from both (12) and (13), $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$, are used as instruments for $y_{1t}$ and $y_{2t}$ when estimating (14). Of course three parameters have been fixed, and so they need to be allowed to vary, and this is done by defining $a_{12}^0 = \frac{\theta_1}{1 - \text{abs}(\theta_1)}, a_{13}^0 = \frac{\theta_2}{1 - \text{abs}(\theta_2)}, a_{23}^0 = \frac{\theta_3}{1 - \text{abs}(\theta_3)}$, and then getting realizations of $\theta_1, \theta_2$ and $\theta_3$ from a uniform random generator. Note that three different random variables $\theta_j$ are needed and these correspond to the three $\lambda_j$ in the Givens matrices.

Unlike the market model it is not easy to find impulse responses that satisfy the sign restrictions. For both methods only around 5% of the impulse responses are retained, and this suggests that the data does not favour the postulated signs for the impulse responses. 1000 of these impulse responses were plotted for SRR in Figure 1 of Fry and Pagan. Therefore, figure 1 below gives the same number of impulse responses from the SRC method (here the positive cost shocks mean a negative productivity shock and, because, Fry

\footnote{For illustration we assume a SVAR or order one, but in the empirical work it is of order two.}

\footnote{Of course since the SVAR is exactly identified this IV procedure is just FIML. The reason for explaining it in terms of IV is that such an approach will be useful when we come to permanent shocks. Nevertheless, given that programs like EViews and Stata estimate the SVARs by FIML it will generally be easier to just set $a_{ij}$ to values and then perform FIML.}
and Pagan used a positive productivity shock in their figure an allowance needs to be made for that when effecting a comparison. It seems as if SRC produces a broader range of impulse responses than SRR, e.g. the maximal contemporaneous effect of demand on output with SRC is more than twice what it is for SRR (we note that all impulse responses in the ranges for both SRC and SRR are valid in that they have the correct signs and they are all observationally equivalent).\(^6\) It is clear that there is a large spread of values i.e. many impulse responses can be found that preserve the sign information and which fit the data equally. The spread here is *across models* and has nothing to do with the variation in data. Hence it is invalid to refer to this range as a "confidence interval" as is often done in the literature. Of course in practice we don’t know \(A_1, \Omega\) and that will make for a confidence interval. We return to the issue in section 8. Such dependence on the data provides some possible extra variation in the spread for impulse responses, but it doesn’t help to conflate this with the variation in them across observationally equivalent models.

6 Combining Parametric and Sign Restrictions

An example is given here featuring the small macro model of section 2 but now with a permanent shock. We compare SRC and SRR and find that the methodologies proceed in the same way. To ensure that there is at least one permanent shock we will assume that the log level of GDP is an \(I(1)\) process and call it \(z_{1t}\.\)\(^7\) Hence \(y_{1t} = \Delta z_{1t}\).

\(^6\)This points to the fact that the impulses found with SRC and SRR may not span the same space. Thinking of this in the context of the market model it is clear that we could find an \(\alpha\) (for SRC) that would exactly reproduce the same \(\alpha\) as coming from SRR. But the estimate of \(\beta\) found by both methods would then differ, and that would lead to different impulse responses. These two sets of impulse responses will be connected by a non-singular transformation but it will vary from trial to trial. If it did not vary then the impulse responses would span the same space.

\(^7\)In Cho and Moreno the output gap was formed by regressing \(z_{1t}\) against a constant and a time trend and then using the residuals to measure it, so the underlying assumption was that \(z_{1t}\) was stationary around a deterministic trend.
Figure 1: 1000 Impulses Responses from SRC Satisfying the Sign Restrictions for the Small Macro Model using the Cho-Moreno Data
The SVAR system composed of $y_{1t}$, $y_{2t}$ and $y_{3t}$ has one permanent (supply) shock in the system plus two transitory shocks associated with demand and an interest rate. By definition these transitory shocks both have a zero long run effect on output, $z_{1t}$. Before imposing any long-run restrictions the SVAR(1) system would be

$$
\Delta z_{1t} = a_{12}^0 y_{2t} + a_{13}^0 y_{3t} + a_{12}^1 y_{2t-1} + a_{13}^1 y_{3t-1} + a_{11}^1 \Delta z_{1t-1} + \varepsilon_{1t} \quad (15)
$$

$$
y_{2t} = a_{21}^0 \Delta z_{1t} + a_{23}^0 y_{3t} + a_{22}^1 y_{2t-1} + a_{23}^1 y_{3t-1} + a_{21}^1 \Delta z_{1t-1} + \varepsilon_{2t} \quad (16)
$$

$$
y_{3t} = a_{31}^0 \Delta z_{1t} + a_{32}^0 y_{2t} + a_{32}^1 y_{2t-1} + a_{33}^1 y_{3t-1} + a_{31}^1 \Delta z_{1t-1} + \varepsilon_{2t} \quad (17)
$$

Now the two transitory shocks must have a zero long-run effect upon output, and we take these to be the second and third ones. Following Fisher et al (2014) this restriction can be imposed on the system (15)-(17) by using the Shapiro and Watson (1988) approach of replacing (15) with

$$
\Delta z_{1t} = a_{12}^0 \Delta y_{2t} + a_{13}^0 \Delta y_{3t} + a_{11}^1 \Delta z_{1t-1} + \varepsilon_{1t},
$$

thereby allowing the parameters of (18) to be estimated by using $y_{2t-1}$, $y_{3t-1}$ and $\Delta z_{1t-1}$ as instruments. Once parameter estimates for (18) are obtained one can get residuals $\hat{\varepsilon}_{1t}$. This is the first step in implementing both the SRC and SRR approaches. Unlike the market model in which a recursive model provided the initial impulse responses to be re-combined in SRR, it is now necessary to use a non-recursive system that incorporates (18) in order to ensure that there will be one permanent and two transitory shocks in the system.

Having recovered the permanent shock $\varepsilon_{1t}$ by the use of parametric restrictions, it will be the case that the permanent impulse responses are known, and they will not change in different trials i.e. they are not re-combined in SRR. To understand why this is so, suppose the SVAR is written as

$$
A_0 \zeta_t = A_1 \zeta_{t-1} + \varepsilon_t,
$$

where $\zeta_t$ is the $3 \times 1$ vector $\begin{bmatrix} \Delta z_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix}$. Then the moving average representation is $\zeta_t = C_0 \varepsilon_t + C_1 \varepsilon_{t-1} + \ldots$. Because $E(\varepsilon_t \varepsilon_{t-j}) = 0$ for $j > 0$, we can recover $C_0$ by regressing $\zeta_t$ against $\varepsilon_t$. Looking at that regression for the first variable it would have the form $\Delta z_{1t} = C_{11}^0 \varepsilon_{1t} + C_{12}^0 \varepsilon_{2t} + C_{13}^0 \varepsilon_{3t}$. Furthermore, because $\varepsilon_{1t}$ is uncorrelated with $\varepsilon_{2t}$ and $\varepsilon_{3t}$, $C_{11}^0$ can be estimated by regressing $\Delta z_{1t}$ on $\varepsilon_{1t}$.\footnote{The same argument applies to regressing the other two variables in $z_t$ to get their impulse responses to the permanent shock $\varepsilon_{1t}$.} By the same argument all that is needed in
order to recover the impulse responses for the permanent shock are $\zeta_t$ (data) and an estimate of $\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t}$. This is an example of the maxim in Fry and Pagan (2011) that one recombines only the base permanent (transitory) shocks when forming the new permanent (transitory) shocks. If this is not done then the recombined shocks will all have permanent effects. Consequently, when there is just a single permanent shock, once it is found there is nothing for it to be combined with to produce new permanent shocks. It is only the transitory shocks that can be re-combined and these will be $\hat{\varepsilon}_{2t}$ and $\hat{\varepsilon}_{3t}$.

SRR will find some base transitory shocks in the following way. Set $a_{23}^0 = 0$ and then estimate (16) using $\hat{\varepsilon}_{1t}$ as an instrument for $\Delta z_{1t}$. Then $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$ can be used to estimate (17) and the shock $\hat{\varepsilon}_{3t}$ follows. The impulse responses for $\hat{\varepsilon}_{2t}$ and $\hat{\varepsilon}_{3t}$ are then recombined to find new transitory shocks. So it is necessary to impose the long run restriction and a recursive assumption to find the initial base shocks.

Now look at the SRC methodology. This requires that the second equation (16) be estimated and $\hat{\varepsilon}_{1t}, y_{2t-1}, y_{3t-1}$, and $\Delta z_{1t-1}$ are available as the instruments for this purpose. But this is one fewer instrument than is needed. To overcome this problem fix $a_{23}^0$ and create a new dependent variable $y_{2t} - a_{23}^0 y_{3t}$. There are now the correct number of instruments and, once the equation is estimated, residuals $\hat{\varepsilon}_{2t}$ would be available. These can be used along with $\hat{\varepsilon}_{1t}, y_{2t-1}, y_{3t-1}$ and $\Delta z_{1t-1}$ to estimate the last equation. Thus the SRC method replaces $a_{23}^0$ with some value, and this is exactly the same situation as occurred with the market model, i.e. once $a_{23}^0$ is replaced by some function of $\theta$, every $\theta$ produces new impulse responses and a set of impulse responses. It is crucial to note however that, as $\theta$ is varied, the long-run restriction is always enforced by design of the SVAR i.e. by using (18) as part of it. Because this parametric (long-run) restriction reduced the number of parameters to be estimated by one, only one parameter needs to be prescribed in order to get all the impulse responses. Sign restrictions are applied to determine which of the two transitory shocks is demand and which is monetary policy. Because the permanent shock does not depend in any way upon the values assigned to $a_{23}^0$, it is invariant to the changing values of this coefficient, and so it remains the same (just as the SRR impulse responses were invariant to

\[9\]There are other parametric restrictions that might be applied apart from long-run ones and these would also generate instruments. Suppose that the second shock is not transitory but is taken to have a zero contemporaneous impact on output. Then the VAR (reduced form) residuals for $\Delta z_{1t}$ can be used as an instrument in the second structural equation. Consequently, zero restrictions upon $C_0$ are easily handled in the SRC methodology.
Estimating the SVAR with a permanent shock by the SRC technique now results in 45% of the responses satisfying all the sign restrictions, as compared to the 5% with purely transitory shocks. Consequently, it seems that the data are more compatible with the sign restrictions, provided one allows for a permanent supply side shock to GDP.

7 Finding The Standard Deviations of Shocks for SRR

The SRR process always starts with some base shocks that are uncorrelated with unit variance. Suppose one started with \( v_{it} = \frac{\varepsilon_{it}}{\sigma_i} \), where \( \varepsilon_{it} \) are the true shocks and \( \sigma_i \) are the true standard deviations. Then the base shocks would be \( \eta_{it} = v_{it} \). If these gave impulses satisfying the sign restrictions a rise of one unit in \( \eta_{it} \) would mean a rise in \( \varepsilon_{it} \) of \( \sigma_i \) i.e. the impulse responses identified by sign restrictions are for one standard deviation changes in the true shocks. The problem is that we don’t know what \( \sigma_i \) is. In terms of the market model the issue is that \( \varepsilon_{Dt} \overset{i.i.d.}{\sim} (0, \sigma^2_D) \), \( \varepsilon_{St} \overset{i.i.d.}{\sim} (0, \sigma^2_S) \) and, by setting \( \eta_{1t} = \sigma^{-1}_S \varepsilon_{St}, \eta_{2t} = \sigma^{-1}_D \varepsilon_{Dt} \), the demand and supply equations have been converted to a structural system that has shocks with a unit variance. What we really want are impulse responses to the demand and supply shocks and not to the \( \eta_{it} \). The latter have the same signs as those for \( \varepsilon_{Dt}, \varepsilon_{St} \) but, because sign information is invariant to the "magnitude" of shocks, one does not directly recover the standard deviation of the shocks of interest. Much of the literature seems to treat the impulse responses as if they were responses to a one unit shock, and this is clearly incorrect.

How then is it that \( \sigma_i \) can be estimated either when parametric restrictions or the SRC method are applied? The answer lies in the normalization used in those methods i.e. the \( A_0 \) matrix has unity on the diagonals. Consequently, a normalization also has to be used to find \( \sigma_i \) after the SRR method has been followed. Now, because \( A_0 = C_0^{-1} \), \( \sigma_i \) can be found by transforming the \( C_0^{-1} \) coming from SRR to a form where it has diagonal elements that are

\[ \lambda \).

10In a recursive system a one standard deviation to a shock can be found by looking at the response of the variable that is the dependent variable of the equation that the shock is attached to. But this is not true in non-recursive systems, and sign restrictions generate many non-recursive systems. Only if the correct model is recursive would we be able to simply infer the standard deviation from the response of a model variable.
unity. Let $c_{ij}^0$ be the elements in $C_0^{-1}$. Then dividing each row of $C_0^{-1}$ with $c_{ii}^0$ will produce a matrix $A_0$ that has unity on the diagonals. The factor $\frac{1}{c_{ii}^0}$ will therefore be the estimate of $\sigma_i$.

To illustrate this procedure take the $C_0$ from SRR that was closest to the true value for the market model i.e. $C_0 = \begin{bmatrix} .2472 & - .3563 \\ .7648 & .3509 \end{bmatrix}$. This gives

$$C_0^{-1} = \begin{bmatrix} .9793 & .9910 \\ -2.1272 & .6876 \end{bmatrix},$$

from which the standard deviations would be $c_{11}^* = \frac{1}{.9793} = 1.02$ (for the demand shock) and $c_{22}^* = \frac{1}{.6876} = 1.45$ (for the supply shock). These compare to the true standard deviations of 1 and $\sqrt{2} = 1.42$. Of course this means that because many impulse responses are produced by SRR (and SRC), there will be many values for $\sigma_i$. Just as impulse responses need to be summarized in some way, this will be equally true of the $\sigma_i^{(m)}$ found at the $m'\text{th}$ trial. There is not just one single standard deviation, unless a particular value for $m$ is chosen by some criterion.

8 Standard Errors for Sign Restricted Impulses

8.1 The SRR Method

Let $\hat{C}_j$ be the impulse responses for one standard deviation shocks from a recursive model. These imply that

$$z_t = \hat{C}_{0} \hat{\eta}_t + \hat{C}_{1} \hat{\eta}_{t-1} + \ldots,$$

where $\hat{\eta}_t$ are the standardized recursive shocks i.e. the base shocks. Then

$$z_t = \hat{C}_0 Q' \eta_t + \hat{C}_1 Q' \eta_{t-1} + \ldots = \hat{C}_0 Q' \eta_t^* + \hat{C}_1 Q' \eta_{t-1}^* + \ldots,$$

and $\eta_t^*$ are the sign-restricted shocks. We therefore have that $\hat{C}_j^* = \hat{C}_j Q'$. Now assume that $\hat{C}_j$ is normal with mean $\bar{C}_j$ and variance $V$ (at least in large samples). The mean need not be equal to the true impulse responses since the recursive model that begins the process is most likely mis-specified.
Hence the mean of $\hat{C}_j$ will be $\bar{C}_jQ'$ while the variance of $\hat{C}_j$ will be

$$\text{var}(\hat{C}_j) = \text{var}(\hat{C}_j - \bar{C}_jQ') = E(\hat{C}_jQ' - \bar{C}_jQ') = \text{var}(\bar{C}_j).$$

Consequently the standard errors of the impulses for any model are the same as those for the standardized recursive model. The bias of course will be different, being $(C_j - \bar{C}_j)Q'$. Nevertheless the square of this will be $(C_j - \bar{C}_j)(C_j - \bar{C}_j)'$, so that the mean square error must be the same across all models. Of course this reflects the fact that they are observationally equivalent.

### 8.2 The SRC method

In the case of the SRC method the standard errors will reflect the method used to capture the estimable parameters. It is possible to use any method that will estimate the parameters of a structural system e.g. FIML, IV, Bayesian methods. The standard errors found will vary from realization to realization i.e. for different $\theta$. Once a model is selected e.g. by using the Median Target method of Fry and Pagan (2011), then standard errors follow immediately.

### 9 Conclusion

The paper has looked at two methods for finding impulse responses using sign restrictions from SVARs. It was argued that the sign restrictions literature involves two steps. In the first stage many impulse responses are generated for uncorrelated shocks. The second stage then determines what these shocks should be named based on the sign restrictions. When all shocks are transitory, and there are no parametric restrictions, the traditional approach to finding many impulse responses has been to re-combine an initial set that have certain properties. Instead we advocated a new approach that works directly with the SVAR contemporaneous matrix, replacing non-estimable quantities in it with randomly chosen values and then estimating the remaining parameters so as to produce uncorrelated shocks. This perspective
has many advantages when there are mixtures of parametric and sign restrictions to be used in SVARs. The new method was shown to be simple to apply and it performs as well as the traditional method. Because it focuses directly on the structural equations it can be applied in any context for which problems are formulated in this way, and enables one to apply a wide range of estimation methods.

10 References


