Towards Understanding the Normalization in Structural VAR Models

Kociecki, Andrzej

National Bank of Poland

17 June 2013

TOWARDS UNDERSTANDING THE NORMALIZATION IN STRUCTURAL VAR MODELS

Andrzej Kocięcki
National Bank of Poland
e–mail: andrzej.kociecki@nbp.pl

First version: November 2012
This version: November 2012

Abstract: The aim of the paper is to study the nature of normalization in Structural VAR models. Noting that normalization is the integral part of identification of a model, we provide a general characterization of the normalization. In consequence some the easy–to–check conditions for a Structural VAR to be normalized are worked out. Extensive comparison between our approach and that of Waggoner and Zha (2003a) is made. Lastly we illustrate our approach with the help of five variables monetary Structural VAR model.

I. INTRODUCTION

The great merit of Waggoner and Zha (2003a) and Hamilton et al (2007) is that they made us realize how subtle the normalization in Structural VAR (SVAR) models is. That is contrary to the common view normalization can influence the small sample probabilistic inference in an uncontrollable and strange way. Poor normalization can lead to multimodal small sample distributions of maximum likelihood estimates or multimodal posterior distributions (of function) of parameters of interest. This phenomenon will manifest itself in a rather inaccurate description of statistical uncertainty.

The reaction to the normalization puzzle in SVAR models was the so–called likelihood preserving (*LP*) normalization proposed by Waggoner and Zha (2003a). The procedure relies on the fact that in a SVAR model, which is identified up to a sign of each equation, the likelihood possesses a multitude of global modes. In order to make the probabilistic statements reliable we should choose one mode and focus on the shape of the likelihood in the appropriate area around this mode (we should not mix "contents" of two or several modes in terms of their statistical uncertainty).

Our position is different. We take seriously the statement of Hamilton et al. (2007) that "*the normalization problem is fundamentally a question of identification*". Although Waggoner and Zha (2003a) motivated and justified their normalization rule in the context of well–behaved impulse response functions (IRF's), our contribution is the remark that ill–behaved IRF's are consequence of the lack of global identification. Building on this we proposed different normalization rule whose primary role is to achieve this global identification. The merit of our normalization rule is that it is mathematically deduced straight from the very basic definition of normalization. On the other hand the *LP* normalization is based on informal reasoning (though it possesses some desirable properties). Moreover our normalization is ordinal. That is we lay down a priori permissible sign for a subset of coefficients in the contemporaneous matrix. In contrast using the *LP* normalization we sometimes multiply coefficients in a given equation by 1 but sometimes by minus 1 (depending on what is closer to the mode).

We offer theoretical insight (which is contrary to the common understanding) that sometimes it is not sufficient to restrict the sign of one coefficient in every equation to normalize a model. As we will show in non–recursive SVAR models we usually need more sign restrictions. In fact this may be also perceived as the significant contribution of our paper since at this level of the SVAR model design the economic theory must enter the scene e.g. you must take a stand *a priori* on whether supply is downward or upward sloping.

It turns out that our normalization also sheds some new light on more and more popular sign restrictions used to identify the impact of some shocks in a SVAR, see e.g. Uhlig (2005). Although it is well understood that using only sign restrictions is insufficient to identify a model we claim that they may be necessary for identification. That is according to our theory sign restrictions may constitute inevitable part of the identification of a model i.e. without them a model may be unidentified even if the conditions given in Rubio–Ramírez et al. (2010) are met.

Since the articles of Waggoner and Zha (2003a) and Hamilton, Waggoner and Zha (2007) are frequently cited in the sequel they will be referred to as WZ and HWZ, respectively.

II. General Definition of Normalization in SVAR Model

The subject of this paper is to understand the nature of normalization in the following SVAR model

$$y_t'A = c + y_{t-1}'A_1 + \cdots + y_{t-p}'A_p + \varepsilon_t' ; \qquad \text{for } t = 1,\ldots,T \tag{1}$$

where $A:(m \times m)$ is the nonsingular matrix of contemporaneous relations between the data $y_t:(m\times 1)$, $A_i:(m\times m)$, $c:(1\times m)$ is a vector of constants and $\varepsilon_t \mid y_{t-1}, y_{t-2},\ldots \sim N(0_{m\times 1}, \mathrm{I}_m)$. Let us define $B' = [c' \; A_1' \ldots A_p']$. Further let O_m denote the space of $(m\times m)$ orthogonal matrices i.e. $O_m = \{g \in \mathbb{R}^{m\times m} \mid g'g = gg' = \mathrm{I}_m\}$.

Assume that restrictions identify a SVAR model up to arbitrary sign of each equation. Let us denote this restricted parameter space as $\Theta_{A,B}^r$. To distinguish the identification up to arbitrary sign of each equation from the concept of global identification we term the former as the regional identification. The label "regional" prompts that this is more than local property. Formally

Definition 1: *The SVAR is regionally identified at* $(A,B)\in\Theta_{A,B}^r$ *if and only if (ifif)*[1] $S_{A,B} = \{g\in O_m \mid (Ag, Bg)\in \Theta_{A,B}^r\} = D$ *, where* $D = \{diag(\delta_1,\ldots,\delta_m) \mid \delta_i = \pm 1\}$.

Following the literature we confine ourselves to the normalization put only on elements of A. Hence we restrict permissible A's to some subset $\Theta_A^n \subset \mathbb{R}^{m\times m}$ which entails inequalities on some entries in A matrix. It just amounts to augmenting $\Theta_{A,B}^r$ with inequality constraints Θ_A^n. Let R be the space of all regionally identified parameter points such that $A\in\Theta_A^n$ i.e. $R = \{(A,B)\in\Theta_{A,B}^r \mid A\in\Theta_A^n$ and (A,B) is regionally identified$\}$. Of course we must assume that $R \neq \varnothing$. Hence we have intuitively clear

Definition 2: *A normalization is a subset* $\Theta_A^n \subset \mathbb{R}^{m\times m}$ *such that for all* $(A,B)\in R$ *we have* $\{g\in O_m \mid (Ag,Bg)\in\Theta_{A,B}^r, Ag\in\Theta_A^n\} = \{\mathrm{I}_m\}$.

[1] We use "ifif" instead of the usual "iff" following suggestion of I.J. Good. He used to say that "ifif" is at least pronounceable neologism ("iff" is the barbarism).

Hence to normalize a model means to achieve global identification on R (uniformly). That is when a normalization is imposed the point that is regionally identified becomes globally identified.

Let $(A,B)\in R$. Since
$\{g\in O_m \mid (Ag,Bg)\in\Theta^r_{A,B}, Ag\in\Theta^n_A\}=\{g\in O_m \mid (Ag,Bg)\in\Theta^r_{A,B}\}\cap\{g\in O_m \mid Ag\in\Theta^n_A\}$
We obtain by definition 1

$$\{g\in O_m \mid (Ag,Bg)\in\Theta^r_{A,B}, Ag\in\Theta^n_A\}=D\cap\{g\in O_m \mid Ag\in\Theta^n_A\}\subseteq D \tag{2}$$

Hence equivalent definition of normalization is

Definition 2A: *A normalization is a subset $\Theta^n_A\subset\mathbb{R}^{m\times m}$ such that for all $(A,B)\in R$ we have $\{g\in D \mid (Ag,Bg)\in\Theta^r_{A,B}, Ag\in\Theta^n_A\}=\{\mathrm{I}_m\}$.*

WZ refer to "conventional" normalization as the sign choice of arbitrary (nonzero) element in each equation. With the help of simple 2–dimensional recursive model they demonstrate that "conventional" normalization may entail apparent ill–behavior of the probability statements for impulse responses. Since this example constitutes the motivation to develop "appropriate" normalization rule by WZ we first analyze this example from our perspective (which is complementary to that adopted by WZ). To this end we ignore any lags in SVAR since they do not play any role in our reasoning. Consider the simple SVAR $y_t'A=\varepsilon_t'$, where $A=\begin{bmatrix}a_{11} & 0\\ a_{21} & a_{22}\end{bmatrix}$ i.e. $\Theta^r_A=\begin{bmatrix}\mathbb{R} & 0\\ \mathbb{R} & \mathbb{R}\end{bmatrix}$. Evidently $\mathrm{I}_2\in\Theta^r_A$ and $S_{\mathrm{I}_2}=\{g\in O_2 \mid \mathrm{I}_2 g\in\Theta^r_A\}=D$. Hence the model is regionally identified at $A=\mathrm{I}_2$. As noted by WZ if we employ the normalization that all diagonal elements in A are positive i.e. $\Theta^n_A=\{A\in\mathbb{R}^{2\times2} \mid a_{11}>0, a_{22}>0\}$, which leads to $R=\{A\in\Theta^r_A \mid A\in\Theta^n_A\}=\Theta^r_A\cap\Theta^n_A=\begin{bmatrix}\mathbb{R}^+ & 0\\ \mathbb{R} & \mathbb{R}^+\end{bmatrix}$, then $\mathrm{I}_2\in\Theta^r_A\cap\Theta^n_A$ and $\{g\in O_2 \mid \mathrm{I}_2 g\in\Theta^r_A\cap\Theta^n_A\}=\{\mathrm{I}_2\}$ i.e. the model is globally identified at $A=\mathrm{I}_2$. Using the "conventional" normalization argument we could replace Θ^n_A with $\bar\Theta^n_A=\{A\in\mathbb{R}^{2\times2} \mid a_{21}>0, a_{22}>0\}$, which induces $R=\{A\in\Theta^r_A \mid A\in\bar\Theta^n_A\}=$ $=\Theta^r_A\cap\bar\Theta^n_A=\begin{bmatrix}\mathbb{R} & 0\\ \mathbb{R}^+ & \mathbb{R}^+\end{bmatrix}$. WZ show how strange results may appear when normalization $\bar\Theta^n_A$ is used. However to point out clear unreasonableness of normalization $\bar\Theta^n_A$ it is sufficient to realize that with this normalization $\mathrm{I}_2\notin R$ i.e. the point that is globally identified using Θ_A^n ceased to be so with $\bar{\Theta}_A^n$. Hence using $\bar{\Theta}_A^n$ we "lose" from potentially normalized (globally identified) set some regionally identified points. The general question is this: Is it reasonable to sacrifice some points from globally identified set when they are economically reasonable? This suggests that the arbitrariness of "conventional" normalization is illusory and must be abandoned. Formal analysis is calling for.

Using definition 2 we have a useful alternative characterization of normalization

Proposition 1: *A normalization is a subset* $\Theta_A^n \subset \mathbb{R}^{m\times m}$ *such that for all* $(A,B)\in R$; $D\cap A^{-1}\Theta_A^n = \{\mathrm{I}_m\}$, *where* $A^{-1}\Theta_A^n = \{A^{-1}\bar{A} \mid \bar{A}\in\Theta_A^n\}$.

Proof: Let $(A,B)\in R$. We have $\{g\in O_m \mid (Ag,Bg)\in\Theta_{A,B}^r, Ag\in\Theta_A^n\} = = D\cap\{g\in O_m \mid Ag\in\Theta_A^n\}$ by (2). We shall prove $\{g\in O_m \mid Ag\in\Theta_A^n\} = = \{g\in O_m \mid g\in A^{-1}\Theta_A^n\}$. We get

$Ag\in\Theta_A^n \;\Rightarrow\; Ag=\bar{A}$ for some $\bar{A}\in\Theta_A^n \;\Leftrightarrow\; g=A^{-1}\bar{A}$ for some $\bar{A}\in\Theta_A^n \Rightarrow g\in A^{-1}\Theta_A^n$

On the other hand

$g\in A^{-1}\Theta_A^n \;\Rightarrow\; g=A^{-1}\bar{A}$ for some $\bar{A}\in\Theta_A^n \;\Leftrightarrow\; Ag=\bar{A}$ for some $\bar{A}\in\Theta_A^n \Rightarrow Ag\in\Theta_A^n$

Then for arbitrary $(A,B)\in R$

$$\{g\in O_m \mid (Ag,Bg)\in\Theta_{A,B}^r, Ag\in\Theta_A^n\} = D\cap\{g\in O_m \mid g\in A^{-1}\Theta_A^n\} = D\cap O_m\cap A^{-1}\Theta_A^n = = D\cap A^{-1}\Theta_A^n$$

The result follows by replacing $\{g\in O_m \mid (Ag,Bg)\in\Theta_{A,B}^r, Ag\in\Theta_A^n\}$ with $D\cap A^{-1}\Theta_A^n$ in definition 2.

Proposition 1 is the most general and valid for all kinds of homogenous restrictions (i.e. linear or nonlinear, whether the restricted parameter space is variation free or not etc.). However if the restricted parameter space is variation free i.e. when restrictions on A and B are independent of each other, then more intuitive and easily checkable condition is available (see below). Proposition 1 says that a model is normalized ifif for all $(A,B)\in R$ the only common element of the two subsets D and $A^{-1}\Theta_A^n$ is the identity matrix. In principle, to achieve this global identification we must have that $\forall(A,B)\in R$, $A^{-1}\Theta_A^n$ is a subset with all diagonal elements strictly greater than minus one (since $D=\{diag(\delta_1,\ldots,\delta_m) \mid \delta_i=\pm1\}$). However even in very special cases this is not feasible e.g. when A is lower triangular and B unrestricted. That is why we are left with other condition. In particular we must ensure that $\forall(A,B)\in R$, $A^{-1}\Theta_A^n$ is a subset with strictly positive diagonal elements. Using proposition 1 we automatically rule out cases when A is singular and/or $A^{-1}\Theta_A^n$ is empty. The very reason for this is that our definition of normalization is uniform i.e. it has to be fulfilled for all $(A,B)\in R$.

In the rest of paper we focus on the case when $\Theta_{A,B}^r=\Theta_A^r\times\Theta_B^r$ i.e. restricted parameter space is variation free so as the restrictions on A and B are independent of each other. This covers unquestionably the most common use of SVAR model in which restrictions are confined to A matrix only. The main reason for that is to have a close contact with WZ (and to be concise).

Anyway though in general case our starting point to study normalization would be proposition 1, when $\Theta_{A,B}^r=\Theta_A^r\times\Theta_B^r$ we have

Proposition 2: *Suppose* $\Theta_{A,B}^r=\Theta_A^r\times\Theta_B^r$. *If* $\forall(A,B)\in R$; $D\cap A^{-1}(\Theta_A^r\cap\Theta_A^n)=\{\mathrm{I}_m\}$ *then* Θ_A^n *is a normalization, where* $A^{-1}(\Theta_A^r\cap\Theta_A^n)=\{A^{-1}\overline{A}\mid\overline{A}\in\Theta_A^r\cap\Theta_A^n\}$.

Proof: First note that given $\Theta_{A,B}^r=\Theta_A^r\times\Theta_B^r$ we have

$\{g\in D\mid(Ag,Bg)\in\Theta_{A,B}^r,Ag\in\Theta_A^n\}=\{g\in D\mid Ag\in\Theta_A^r\cap\Theta_A^n,Bg\in\Theta_B^r\}=$

$=D\cap\{g\in D\mid Ag\in\Theta_A^r\cap\Theta_A^n\}\cap\{g\in D\mid Bg\in\Theta_B^r\}$

Analogously as in the proof of proposition 1 we can show $\{g\in D\mid Ag\in\Theta_A^r\cap\Theta_A^n\}$ $=\{g\in D\mid g\in A^{-1}(\Theta_A^r\cap\Theta_A^n)\}$. Hence

$\{g\in D\mid(Ag,Bg)\in\Theta_{A,B}^r,Ag\in\Theta_A^n\}=D\cap A^{-1}(\Theta_A^r\cap\Theta_A^n)\cap\{g\in D\mid Bg\in\Theta_B^r\}$

Note that if $\Theta_{A,B}^r=\Theta_A^r\times\Theta_B^r$ then $R=\{A\in\Theta_A^r\cap\Theta_A^n,B\in\Theta_B^r\mid(A,B)$ is regionally identified$\}$. Let $A,B\in R$ be arbitrary. It follows

$\{g\in D\mid(Ag,Bg)\in\Theta_{A,B}^r,Ag\in\Theta_A^n\}=D\cap A^{-1}(\Theta_A^r\cap\Theta_A^n)\cap\{g\in D\mid Bg\in\Theta_B^r\}$

If $D\cap A^{-1}(\Theta_A^r\cap\Theta_A^n)=\{\mathrm{I}_m\}$ then $\{g\in D\mid(Ag,Bg)\in\Theta_{A,B}^r,Ag\in\Theta_A^n\}=\{\mathrm{I}_m\}$ since $\{g\in D\mid Bg\in\Theta_B^r\}$ is non–empty and contains I_m. This proves that Θ_A^n is a normalization.

Proposition 2 gives sufficient condition for normalization when $\Theta_{A,B}^r=\Theta_A^r\times\Theta_B^r$ and there are some restrictions imposed on B. However if the restrictions are confined to A only, then proposition 2 constitutes necessary and sufficient condition for normalization. For future reference let us denote $\Theta_A^{rn}\equiv\Theta_A^r\cap\Theta_A^n$.

A normalization consistent with proposition 2 will be refereed to as *PL* normalization (you can think of *PL* as an abbreviation for “plain” or alternatively as the ISO country code for certain medium–sized country in Europe). *PL* normalization is appropriate when restricted parameter space is variation free and requires that for all $(A,B)\in R$, the only common element from two subsets D and $A^{-1}\Theta_A^{rn}$ is the identity matrix. This will be achieved by the condition that all elements in $A^{-1}\Theta_A^{rn}$ have strictly positive diagonal elements. In the important special case when all restrictions are confined to A only we must ensure that for all $A\in\Theta_A^{rn}$ such that A is regionally identified, $A^{-1}\Theta_A^{rn}$ has strictly positive diagonal elements. Or simply for every $A,\overline{A}\in\Theta_A^{rn}$, $A^{-1}\overline{A}$ has strictly positive diagonal elements.

The inevitable question from which we could not escape is how *PL* normalization looks like when using recursive SVAR models i.e. A is lower or upper triangular and B unrestricted. It is easy to note that restricting all elements on the diagonal of A to be positive or negative works well. Let us show this. Without loss of generality assume A is lower triangular with positive diagonal elements. Then A^{-1} is lower triangular with positive diagonal elements too. If we postmultiply A^{-1} by any $\overline{A}$, which is also lower triangular with positive diagonal elements, then a product $A^{-1}\overline{A}$ is lower triangular with positive diagonal elements. Lastly the only common element of D and the space of lower triangular matrices with positive elements is the identity matrix.

III. First Illustration of PL Normalization

As a first illustration of our approach we use the example of orange demand and supply discussed in detail in HWZ. Let $y_t'=(q_t,p_t,w_t)$, where q_t denotes the log of the number of oranges sold in year *t*, p_t is the log of the price and w_t is the number of days with below–freezing temperatures in year *t*. The model is as follows

$$y_t'\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix} = c + y_{t-1}'A_1 + \cdots + y_{t-p}'A_p + \varepsilon_t' \tag{3}$$

where A_i's are unrestricted. Of course the restricted parameter space is variation free hence proposition 2 applies. First equation represents a supply, second – a demand and the last one depicts the exogenous process w_t. One may show that the model (3) is regionally identified almost everywhere [Lebesgue] (in fact provided that $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is nonsingular and $a_{31}\neq 0$). By assumption A must be also non–singular hence $a_{33}\neq 0$. Then

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix} \Rightarrow A^{-1} = [\det(A)]^{-1} \cdot \begin{bmatrix} a_{22}a_{33} & -a_{12}a_{33} & 0 \\ -a_{21}a_{33} & a_{11}a_{33} & 0 \\ -a_{31}a_{22} & a_{31}a_{12} & a_{11}a_{22} - a_{21}a_{12} \end{bmatrix}$$

where $\det(A) = a_{11}a_{22}a_{33} - a_{33}a_{12}a_{21}$.

Let $\overline{A} = \begin{bmatrix} \overline{a}_{11} & \overline{a}_{12} & 0 \\ \overline{a}_{21} & \overline{a}_{22} & 0 \\ \overline{a}_{31} & 0 & \overline{a}_{33} \end{bmatrix}$ be arbitrary. Then diagonal elements of $A^{-1}\overline{A}$ are given as

$$\begin{aligned} d_{11} &= (a_{22}a_{33}\overline{a}_{11} - a_{12}a_{33}\overline{a}_{21})/(a_{11}a_{22}a_{33} - a_{33}a_{12}a_{21}) \\ d_{22} &= (a_{11}a_{33}\overline{a}_{22} - a_{21}a_{33}\overline{a}_{12})/(a_{11}a_{22}a_{33} - a_{33}a_{12}a_{21}) \\ d_{33} &= (a_{11}a_{22}\overline{a}_{33} - a_{21}a_{12}\overline{a}_{33})/(a_{11}a_{22}a_{33} - a_{33}a_{12}a_{21}) \end{aligned} \tag{4}$$

Our goal is to guess Θ_A^n such that for every $A, \overline{A} \in \Theta_A^{rn}$, $A^{-1}\overline{A}$ has strictly positive diagonal elements i.e. $d_{ii} > 0$; $\forall i$. It is easily verified that taking Θ_A^n to be the space of A's with strictly positive diagonal elements is insufficient to complete the global identification of the model. We need one more assumption: a_{12} is positive and a_{21} is negative (or vice versa). Interestingly this requires economic reasoning. Since the first equation is a supply and the second is a demand it is natural to assume $a_{21} < 0$ and $a_{12} > 0$ (note that all coefficients of each contemporaneous relation are on the left side in SVAR (1))[2]. Hence

$$\Theta_A^n = \{A \in \mathbb{R}^{m\times m} \mid a_{11} > 0, a_{22} > 0, a_{33} > 0, a_{21} < 0, a_{12} > 0\} \tag{5}$$

Imposing *PL* normalization allows us to postmultiply (3) by $diag(a_{11}^{-1}, a_{22}^{-1}, a_{33}^{-1})$ to get (with some abuse of notation)

$$y_t' \begin{bmatrix} 1 & \eta & 0 \\ \gamma & 1 & 0 \\ h & 0 & 1 \end{bmatrix} = c + y_{t-1}'A_1 + \cdots + y_{t-p}'A_p + u_t' \tag{6}$$

Note that with *PL* normalization $\eta = a_{12}/a_{22} > 0$, $\gamma = a_{21}/a_{11} < 0$, $h = a_{31}/a_{11}$ and $u_t \sim N(0_{3\times1}, diag(a_{11}^{-2}, a_{22}^{-2}, a_{33}^{-2}))$. The notation used in (6) is not accidental. The specification (6) "almost" corresponds to the so–called $\eta-$normalization in HWZ.

[2] This for example precludes downward sloping supply curve. We are aware that sometimes such a description of the market phenomenon is quite reasonable. The point is that complete identification of the SVAR model requires that you must take a stand *a priori* on whether supply is downward or upward sloping.

We wrote "almost" since $\eta-$normalization in HWZ does not take into account the sign restrictions $\eta > 0$ and $\gamma < 0$, which are in fact necessary for global identification. The interesting fact about (6) under the *PL* normalization is that it roughly conforms to the so–called identification principle i.e. normalization rule proposed by HWZ. Identification principle applied to (6) says that boundaries for allowable entries of A matrix should correspond to the loci along which the log likelihood is $-\infty$. In our case this locus is $\gamma\eta = 1$[3]. Although the locus of $\gamma\eta = 1$ is not on the boundary of the parameter space in (6) what *PL* normalization does instead is to exclude the parameter points for which $\gamma\eta = 1$ holds (since $\eta > 0$ and $\gamma < 0$).

To illustrate all these issues we simulated the sample of 100 observations from

$$y_t' \begin{bmatrix} 1 & 0.1 & 0 \\ -0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.12 \\ 0.3 \end{bmatrix}' + y_{t-1}' \begin{bmatrix} 0.5 & 0 & -0.1 \\ 0.3 & 1 & 0.1 \\ 0 & 0.1 & 0.4 \end{bmatrix} + \varepsilon_t' ; \qquad \varepsilon_t \sim N(0_{3\times 1}, \mathrm{I}_3) \qquad (7)$$

Since $a_{ii} = 1$ for $i = 1,2,3$; we have $a_{12} \equiv \eta = 0.1$, $a_{21} \equiv \gamma = -0.5$ and $a_{31} \equiv h = 0.5$. The first picture in Figure 1 shows the contours of concentrated log likelihood for η and γ evaluated at the true values for $h = 0.5$ and $a_{ii} = 1$, for $i = 1,2,3$. The true values of η and γ are marked with "$\times$". The likelihood has one global maximum and two local peaks (of unequal height). The areas of great concentration of the contour levels correspond to the loci along $\gamma\eta = 1$ (at which the log likelihood approaches $-\infty$). Using the *PL* normalization we restrict the support to the second quadrant i.e. $\eta > 0$ and $\gamma < 0$. Thus we automatically exclude parameters along and in the vicinity of $\gamma\eta = 1$ that are situated both in the first and the third quadrant. It is instructive to find out how these problems carry over into the posterior results. Although the multimodality naturally characterizes the marginal posterior of η and γ (derived under the flat prior for all parameters in SVAR), see the last picture in Figure 1, the contribution of these modes to the visible shape of the posterior is none, see the middle picture in Figure 1. The reason is that the ratio of the height of the marginal posterior of η and γ at the global maximum to that of the second largest peak (around $\gamma = -2$, $\eta = -1.2$) is about e^{150}. In consequence the IRF's computed from (6) even without inequality constraints $\eta > 0$ and $\gamma < 0$ are well behaved too

[3] Which is also the locus of local non–identification that arises at $h = 0$.

i.e. the error bands for IRF's are not too wide and quite conclusive, see HWZ. However in contrast to HWZ we interpret these results differently. For HWZ the parameterization (6) without inequality restrictions $\eta > 0$ and $\gamma < 0$ is acceptable since "*for practical purposes it is sufficiently close* [...] *to a true identification–based normalization*". As we will show in section VI, this conclusion is case–sensitive. In this particular 3–dimensional SVAR subject to the particular identifying scheme, restricting the diagonal elements in A results in well behaved posterior of parameters and its functions e.g. IRF's. In general this is not a rule. In fact this is the message from WZ. Using the PL normalization guarantees well behaved posteriors of parameter and its functions in larger models when the simple visual inspection of the shapes of the likelihood and/or posterior is not readily available and ad–hoc normalization rules are not an option.

Quite obviously those inequality constraints turn out to be also sign restrictions for impulse responses. For example instantaneous response of the price to a one standard deviation positive shock to a quantity supplied is $\varphi_{12} = -a_{12}a_{33}/(a_{11}a_{22}a_{33} - a_{33}a_{12}a_{21})$. Using PL normalization we have $\varphi_{12} < 0$. Moreover the instantaneous effect of a one standard deviation increase in quantity demanded on the price is strictly positive under PL normalization (since $\varphi_{22} = a_{11}a_{33}/(a_{11}a_{22}a_{33} - a_{33}a_{12}a_{21}) > 0$). This ensures that we avoid all pitfalls connected with "conventional" normalizing rules which were convincingly illustrated in Figure 4 in HWZ.

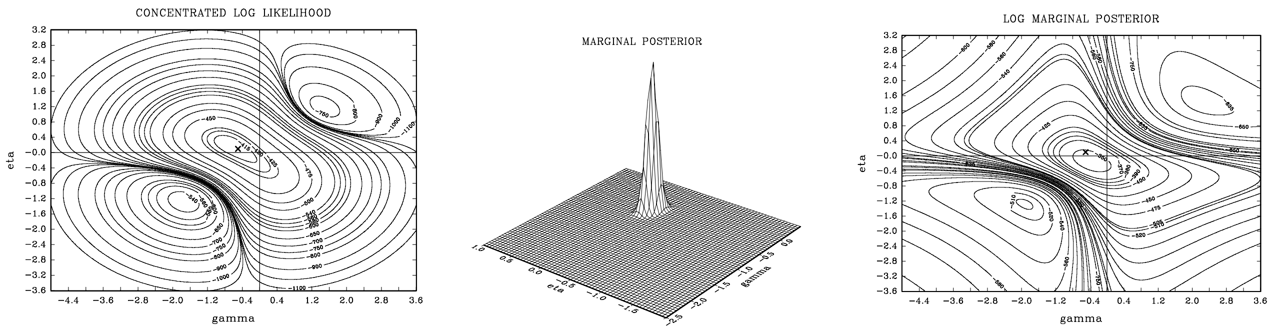

Figure 1: From the left: a) contours of concentrated log likelihood of γ and η evaluated at the true value of $h = 0.5$ and the diagonal elements in A matrix i.e. $a_{ii} = 1$ for $i = 1,2,3$; b) marginal posterior of γ and η under the flat prior for all parameters in SVAR; c) contours of the log marginal posterior of γ and η under the flat prior for all parameters in SVAR.

IV. Sufficient Condition for PL Normalization

In our example of the orange demand–supply, Θ_A^n turns out to be equivalent to assumption that $\forall A \in \Theta_A^{rn}$; $\det(A) > 0$. Was it a coincidence? We will show that provided that the restricted parameter space is variation free this requirement is all we need to complete the global identification of SVAR model.

Among basic model assumptions is that $\det(A) \neq 0$. Commonly this assumption was thought as unimportant because the set of singular matrices has zero Lebesgue measure. For instance using Bayesian simulation methods to estimate SVAR, in practice we could not encounter a draw which entails singular A. However the theoretical importance of the singularity of A has been recognized and discussed by WZ and HWZ. The message from these both articles is that the permitted parameter space should exclude the subspace on which the likelihood vanishes. The goal of this section is to demonstrate that the latter informal statement can be formally justified.

To proceed further we need one more notation. Given two matrices $X = (x_{ij})$ and $Y = (y_{ij})$ of the same dimension we write $X \leq_c Y$ if $x_{ij} \leq y_{ij}$ for each i, j. Hence "$\leq_c$" denotes component–wise inequality. We have

Proposition 3: *Assume* $\Theta^r_{A,B} = \Theta^r_A \times \Theta^r_B$. *Let* $\underline{A}$ *and* $\overline{A}$ *be given matrices in* Θ^r_A. *Assume that* $\Theta^{rn}_A = \{A \in \Theta^r_A \mid \underline{A} \leq_c A \leq_c \overline{A}\}$ *is such that each* $A \in \Theta^{rn}_A$ *is nonsingular. Then for every* $A \in \Theta^{rn}_A$, $A^{-1}\Theta^{rn}_A$ *is a subset with strictly positive diagonal elements.*

Proof: This is the application of theorem 1.2 in Rohn (1989), which states that under hypothesis of our proposition, $A_* = A_1^{-1}A_2$, for every $A_1, A_2 \in \Theta^{rn}_A$, is the so–called $P-$matrix (a square matrix A_* is the $P-$matrix if all its principal minors are positive). In particular since each diagonal element is the principal minor, the proposition follows.

Needless to say some entries in $\overline{A}$ may be set to ∞ and that in $\underline{A}$ to minus ∞.

Proposition 3 forms a basis for useful sufficient condition to achieve global identification. One has to derive analytically $\det(A)$. If it happens that imposing inequalities on some or all entries in A restricts the latter so as $\det(A) > 0$ or $\det(A) < 0$ then setting $\Theta^{rn}_A = \{A \in \Theta^r_A \mid \det(A) > 0\}$ or $\Theta^{rn}_A = \{A \in \Theta^r_A \mid \det(A) < 0\}$ will complete the identification of a model. The choice between $\Theta^n_A = \{A \in \Theta_A \mid \det(A) > 0\}$ and $\Theta^n_A = \{A \in \Theta_A \mid \det(A) < 0\}$ depends on the model at hand (but is only illusory, see right below). Thus instead of finding Θ^n_A such that $\forall A \in \Theta^{rn}_A$; $A^{-1}\Theta^{rn}_A$ is a subset with strictly positive diagonal elements all we have to do is 1) to derive the determinant of A and 2) to impose inequalities on some elements of A in the form $\underline{A} \leq_c A \leq_c \overline{A}$ so as $\det(A) > 0$ or $\det(A) < 0$. Note that derivation of $\det(A)$ even in large SVAR model is usually not very difficult. This is because of many zero restrictions imposed on A. In fact prior to derivation of $\det(A)$, we can permute the rows and columns of A so as there appear blocks of zeros (which usually simplify derivation of $\det(A)$). The permutation operation is permissible since it only changes the sign of the determinant but both $\det(A) > 0$ and $\det(A) < 0$ restriction is acceptable.

Note that when A is lower or upper triangular (or its subset) then if diagonal elements of A are restricted to be positive we immediately get $\det(A) > 0$. This justifies our conviction that the correct normalization is the same for recursive and non–recursive model provided that we follow the rule to restrict A so as $\det(A) > 0$.

V. Comparison of PL Normalization to LP Normalization

It is instructive and desirable to compare our theory with the *LP* normalization proposed by WZ. The first step to apply *LP* normalization is the derivation of the maximum likelihood (ML) estimator of A to be denoted as $\hat{A}$.

Proposition 4: *Assume that* $\Theta^{r}_{A,B} = \Theta^{r}_{A} \times \Theta^{r}_{B}$. *Suppose there is a mode* $\hat{A}$ *in* Θ^{rn}_{A}. *Then the PL normalization implies the LP normalization restricted to* $\Theta^{rn}_{A} \times \Theta^{r}_{B}$.

Proof: *PL* normalization implies that for every two distinct $A, \overline{A} \in \Theta^{rn}_{A}$, $A^{-1}\overline{A}$ must have strictly positive diagonal elements. Since a mode $\hat{A}$ belongs to Θ^{rn}_{A} it follows that for every $A \in \Theta^{rn}_{A}$, $A^{-1}\hat{A}$ has also strictly positive diagonal elements. The latter is the *LP* normalization.

Otherwise, if we operate on $\Theta^{r}_{A,B} = \Theta^{r}_{A} \times \Theta^{r}_{B}$ (and not on $\Theta^{rn}_{A} \times \Theta^{r}_{B}$), the *LP* and *PL* normalizations are incomparable notions. However if $\hat{A} \in \Theta^{rn}_{A}$, by proposition 4, any $A \in \Theta^{rn}_{A}$ is consistent with the *LP* normalization. On the other hand the parameter points that are chosen using the *LP* normalization may not belong to Θ^{rn}_{A} [4]. Thus as a general principle we may expect that error bands for IRF's in a model under *PL* normalization will be narrower than those in a model under the *LP* normalization.

[4] Consider the model of the orange demand–supply from section III. Let us focus on the first diagonal element of $A^{-1}\hat{A}$. Suppose $a_{ii} = 1$; $\forall i$, $a_{12} = 0.1$ and $a_{21} = 0.2$ (note that a_{21} violates the *PL* normalization). Further suppose that ML estimators are $\hat{a}_{11} = 1$ and $\hat{a}_{21} = -0.1$. Then the first diagonal element of $A^{-1}\hat{A}$ is $(1 + 0.1 \cdot 0.1)/(1 - 0.1 \cdot 0.2) > 0$. It follows that $a_{21} = 0.2$ is consistent with the *LP* normalization.

Utilizing the *PL* normalization we may hope for more clear–cut economic conclusions as far as IRF's are concerned. That this hope is justified will be illustrated in section VI.

Without loss of generality assume $\Theta_A^{rn} = \{A \in \Theta_A^r \mid \det(A) > 0\}$. Then we have the following

Proposition 5: *Assume that* $\Theta_{A,B}^r = \Theta_A^r \times \Theta_B^r$. *Suppose* $\Theta_A^{rn} = \{A \in \Theta_A^r \mid \det(A) > 0\}$ *entails the inequalities so as proposition 3 holds. Let* $\Gamma = diag(\gamma_{11}, \dots, \gamma_{mm})$ *be any matrix with* $\gamma_{ii} \in [0,1]$. *Then for all* $\bar{A}, A \in \Theta_A^{rn}$ *we have* $\det(\bar{A}\Gamma + A(\mathrm{I}_m - \Gamma)) > 0$.

Proof: By assumption $\det(\bar{A}\Gamma + A(\mathrm{I}_m - \Gamma)) = \det(A) \cdot \det((\mathrm{I}_m - \Gamma) + A^{-1}\bar{A}\Gamma)$. Then $\det(\bar{A}\Gamma + A(\mathrm{I}_m - \Gamma)) > 0$ ifif $\det((\mathrm{I}_m - \Gamma) + A^{-1}\bar{A}\Gamma) > 0$. By theorem 1.2 in Rohn (1989), for all $\bar{A}, A \in \Theta_A^{rn}$, $A^{-1}\bar{A}$ is a $P-$matrix (hence all principal minors of $A^{-1}\bar{A}$ are positive). Proposition follows by expansion of $\det((\mathrm{I}_m - \Gamma) + A^{-1}\bar{A}\Gamma)$ by the diagonal $(\mathrm{I}_m - \Gamma)$ (see e.g. Seber (2008), pp. 61–62 or Harville (1997), p. 196) and noting that for every Γ and $\bar{A}, A \in \Theta_A^{rn}$, the determinant is positive.

In particular under hypothesis of proposition 5 and provided that $\hat{A} \in \Theta_A^{rn}$ we get $\det(\hat{A}\Gamma + A(\mathrm{I}_m - \Gamma)) = \det([\gamma_{11}\hat{a}_1 + (1-\gamma_{11})a_1, \dots, \gamma_{mm}\hat{a}_m + (1-\gamma_{mm})a_m]) > 0$ for all $\gamma_{ii} \in [0,1]$, where $\hat{a}_i$ denotes the $i-$th column of $\hat{A}$ and a_i that of A. In contrast the *LP* normalization works column–wise so as given $a_1 \dots a_{i-1}, a_{i+1} \dots a_m$, we choose a_i such that $\det([a_1, \dots, a_{i-1}, \gamma\hat{a}_i + (1-\gamma)a_i, a_{i+1}, \dots, a_m]) > 0$, for all $\gamma \in [0,1]$, i.e. $\hat{a}_i$ and a_i lie on the same side of the hyperplane $\{a_i \in \mathbb{R}^m \mid \det(A) = 0,$ given $a_1 \dots a_{i-1}, a_{i+1} \dots a_m\}$. Evidently the *PL* normalization ensures that $\hat{a}_i$ and a_i lie on the same side of the hyperplane but simultaneously for all $i = 1, \dots, m$ and unconditionally (i.e. without conditioning on $a_1 \dots a_{i-1}, a_{i+1} \dots a_m$).

VI. A Monetary Policy Example

As a second (real–data) example we consider a monetary SVAR proposed by Kim (1999). The contemporaneous matrix A is restricted as follows (B unrestricted)

$$A = \begin{array}{r} \\ R \\ \log M \\ \log P \\ \log Y \\ \log P_c \end{array} \begin{array}{c} \begin{array}{ccccc} \text{MP} & \text{MD} & \text{PS} & \text{PS} & \text{Inf} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & 0 & 0 & a_{15} \\ a_{21} & a_{22} & 0 & 0 & a_{25} \\ 0 & a_{32} & a_{33} & 0 & a_{35} \\ 0 & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & 0 & 0 & 0 & a_{55} \end{bmatrix} \end{array} \qquad (8)$$

The identifying scheme is quite similar to that used in Waggoner and Zha (2003b) (except that (8) does not include unemployment). The model includes 5 variables: the federal funds rate (R), logarithm of the monetary aggregate M2 ($\log M$), logarithm of the consumer price index ($\log P$), logarithm of the real GDP interpolated on a monthly frequency and logarithm of the Commodity Research Bureau price index for raw industrial commodities ($\log P_c$). Each column in (8) represents a behavioral equation that is signified at the top. "MP" stands for monetary policy (or money supply) equation and "MD" stands for money demand. Equations labeled with "PS" succinctly describe production sector and "Inf" stands for the information market.

We employed the US dataset used in Waggoner and Zha (2003b), which is available at http://www.tzha.net/computercode (where detailed description of the data may be also found). The data are monthly and cover 01.1959–12.2000.

By theorem 7 in Rubio–Ramírez et al. (2010) the model under the identifying scheme (8) is exactly identified i.e. globally identified almost everywhere. In our language it means that we can define the set of regionally identified parameter points. Since the restricted parameter space is variation free, to get complete identification we need to impose the PL normalization. To this end we use proposition 3 and we have to derive the determinant of A. We easily get

$$\det(A) = a_* a_{44} a_{55} - a_{44}(a_{51}a_{22}a_{33}a_{15} - a_{51}a_{12}a_{33}a_{25})$$

where $a_* = a_{11}a_{22}a_{33} - a_{33}a_{21}a_{12}$. Kim (1999) implicitly assumed that all diagonal elements in A are strictly positive. However it is not sufficient for global identification (i.e. PL normalization). We need to find inequalities on the parameters so as $\det(A) > 0$. Hence in addition to $a_{ii} > 0$ ($i = 1,\dots,5$), we must assume that $a_{21}a_{12} < 0$, $a_{51}a_{15} < 0$ and $a_{51}a_{12}a_{25} > 0$. To use proposition 3 we must derive from the last three inequalities the implied inequalities for single elements. This should be in principle assisted by the economic theory. Using standard economic reasoning since the second equation is money demand we assume $a_{12} > 0$ (by assumption $a_{22} > 0$ and we use a convention that all contemporaneous variables are on the left in SVAR). Moreover $a_{21} < 0$ is also reasonable since we can think of the first equation as money supply (note that we also assume $a_{11} > 0$). Since P_c is the commodity price index in dollars we should expect that monetary authority should increase the interest rates when world commodity price rises. Hence $a_{51} < 0$. Then we must assume $a_{15} > 0$ to fulfill $a_{51}a_{15} < 0$. With these choices we must also restrict $a_{25} < 0$ (so as $a_{51}a_{12}a_{25} > 0$). To rationalize $a_{15} > 0$ and $a_{25} < 0$, Kim (1999) treated the last equation as "*an arbitrage equation which describes a kind of financial market equilibrium*". Thus in case of large economy (like US) domestic interest rates and money aggregates may affect P_c through the direct pressure on world commodity price. When the interest rate rises in large economy it should have tendency towards lowering the commodity prices. That is $a_{15} > 0$. Moreover when the money aggregate increases in large economy there is a natural pressure to increase the commodity prices, hence $a_{25} < 0$. Needless to say the above inequality restrictions put on A give rise to sign restrictions for IRF's (at least for immediate responses).

We estimated the model with $p = 12$ lags using Bayesian approach. To this end we used the flat prior for A and B in order to preserve the likelihood shape. Figures 2 and 3 present IRF's of all variables to a one standard deviation contractionary monetary policy shock and a money demand shock, respectively. These are shocks identified with the first two equations. The solid line presents the IRF evaluated at the maximum likelihood (ML) estimators of A and B. The dashed line (usually very close to the solid one) is median response, two "dots–dashes" lines cover 68% of the posterior probability (point–wise). Lastly two dotted lines are 90% posterior probability bands. In each figure a panel A shows results using naïve normalization i.e. diagonal elements in (8) are positive, panel B demonstrates the output using *LP* normalization and panel C – the *PL* normalization.

The ML estimates imply IRF's expected by economists. For instance, the interest rate rises and a money falls initially, the real GDP declines quite quickly reaching the minimum within half a year and consumer prices decline persistently. Since error bands are meant to describe uncertainty around "mean" response, the probabilistic conclusion may not be so certain and depends on how you normalize the model. With naïve normalization the matter is hopeless which was already nicely demonstrated in WZ. We found out nothing about the most important aspects of the monetary policy shock i.e. its impact on interest rate, real GDP and consumer prices. In general probability bands of all IRF's are suspiciously wide. As we emphasized this is a consequence of the methodological fault and not the "uninformativeness" of the data. Adoption of the *LP* normalization results in more conclusive probability statements for IRF's. However the question of great importance is how *LP* and *PL* normalizations differ from each other and whether these differences are economically important. Firstly using the *PL* normalization we get particularly well determined IRF of the interest rate to a monetary policy shock. We are quite certain that the immediate impact is positive and becomes negative after, say eight months (in anticipation of falling prices and by realizing by monetary policy decision makers that output has already declined). In contrast the analogous IRF using the *LP* normalization gives an ambiguous impression[5]. Secondly with the *PL* normalization we definitely get rid of the "price puzzle" i.e. prices move up after a contractionary monetary policy shock. In this respect probabilistic conclusions are much sharper with *PL* normalization than *LP* normalization.

In fact in all cases *PL* normalization makes the probabilistic statement more informative than when employing the *LP* normalization. Sometimes it is not a critical difference but sometimes it is economically crucial (e.g. compare the response of consumer prices to a money demand shock).

VII. Conclusion

Our goal was to properly grasp the notion of normalization in SVAR models. Using basic definition of normalization in SVAR models we proposed the easy working condition for normalization in SVAR models when the restricted parameter space is variation free. It was called the *PL* normalization. We emphasized that normalization is an integral part of the identification of SVAR. To put it another way, only properly normalized parameter point becomes globally identified.

We compared our theory to the likelihood preserving (*LP*) normalization proposed by Waggoner and Zha (2003a). Our basic attitudes to normalization are quite different. We maintain that a correct approach is to trace the overall shape of the likelihood of the globally identified model whereas Waggoner and Zha (2003a) focus on its shape in the close area around the mode in a model which is "almost" identified (up to arbitrary sign of each equation). In our opinion a proper normalization is not a matter of appropriate description of uncertainty around the maximum likelihood estimate of IRF (as suggested by Waggoner and Zha (2003a)), but is the last and necessary step towards achieving the global identification of SVAR model.

However our theoretical findings are in line with the recommendation of Waggoner and Zha (2003a) and Hamilton et al. (2007) to put the parameter points which imply the zero likelihood or failure of local identification on the boundary of

[5] As a matter of fact the immediate response of the interest rate to a contractionary monetary policy shock is more reasonable when using naïve normalization than the *LP* normalization. In the former case it is "probably" positive.

the parameter space. On the other hand we disagree with Waggoner and Zha (2003a) claim that "*the correct normalization for recursive models turns out to be, in general, inappropriate for nonrecursive models*". Our conclusion is that the correct normalization is the same for recursive and non–recursive SVAR provided that we fully understand what the normalization is (the appropriate normalization rule is the same).

We demonstrated theoretically and in practice that using *PL* normalization we get narrower IRF's error bands than when employing the *LP* normalization. Hence the *PL* normalization may be welcomed by applied macroeconomists as it will tend to confirm more firmly their intuition.

Although general nonlinear identifying restrictions may make the normalization irrelevant (see e.g. Waggoner and Zha (2003a)) there is an important class of nonlinear restrictions (e.g. short–run and long–run impulse response restrictions) that also require normalization rule. Though we provided useful characterization of normalization in such a case (see proposition 1) we really did not study it and leave these aspects of normalization for future research.

REFERENCES:

Hamilton, J.D., D.F. Waggoner and T. Zha (2007), "Normalization in Econometrics", *Econometric Reviews*, 26, pp. 221–252.

Harville, D.A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer–Verlag, New York.

Kim, S. (1999), "Do Monetary Policy Shocks Matter in the G–7 Countries? Using Common Identifying Assumptions about Monetary Policy Across Countries", *Journal of International Economics*, 48, pp. 387–412.

Rohn, J. (1989), "Systems of Linear Interval Equations", *Linear Algebra and Its Applications*, 126, pp. 39–78.

Rubio–Ramírez, J.F, D.F. Waggoner and T. Zha (2010), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference", *The Review of Economic Studies*, 77, pp. 665–696.

Seber, G.A.F (2008), *A Matrix Handbook for Statisticians*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Uhlig, H. (2005), "What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, pp. 381–419.

Waggoner, D.F., and T. Zha (2003a), "Likelihood Preserving Normalization in Multiple Equation Models", *Journal of Econometrics*, 114, pp. 329–347.

Waggoner, D.F., and T. Zha (2003b), "A Gibbs Sampler for Structural Vector Autoregressions", *Journal of Economic Dynamics and Control*, 28, pp. 349–366.

Figure 2 A: Naive normalization i.e. diagonal elements of A matrix are positive

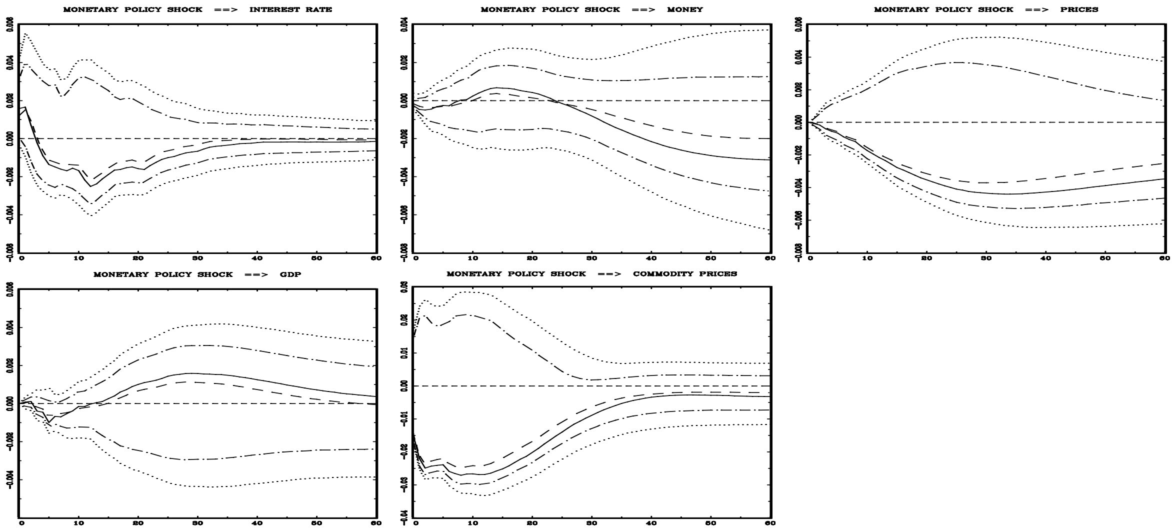

Figure 2 B: LP normalization

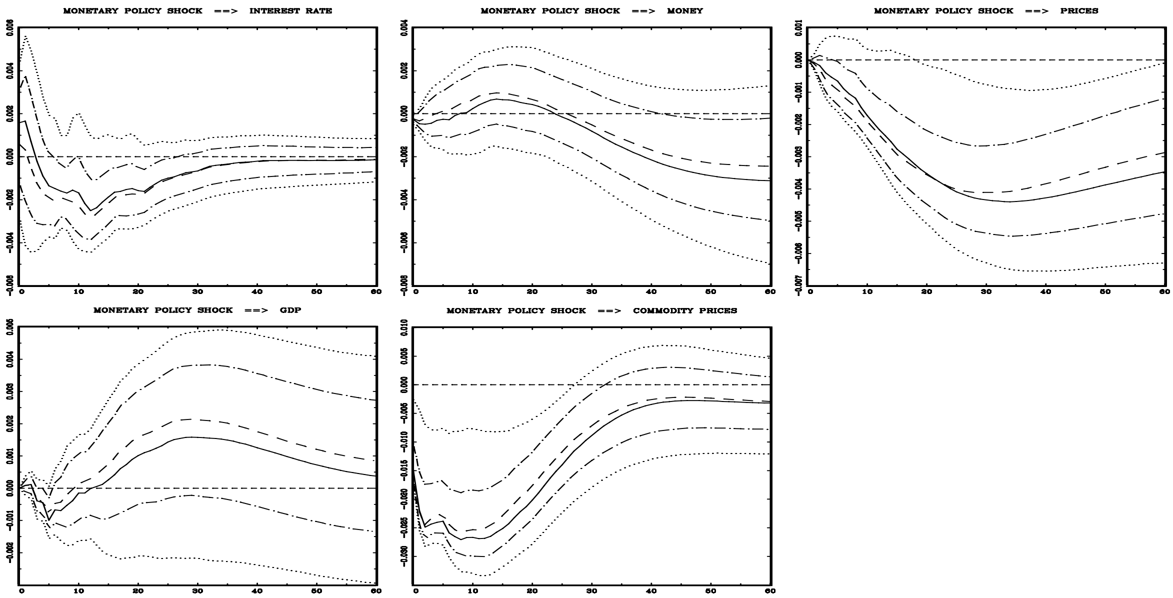

Figure 2 C: PL normalization

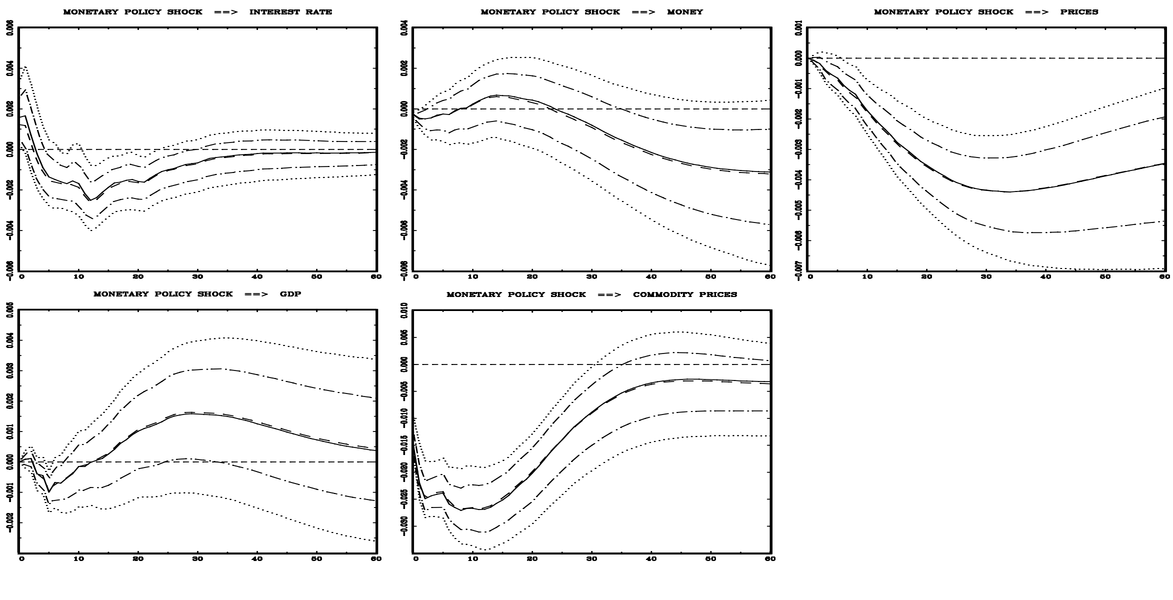

Figure 3 A: Naive normalization i.e. diagonal elements of A matrix are positive

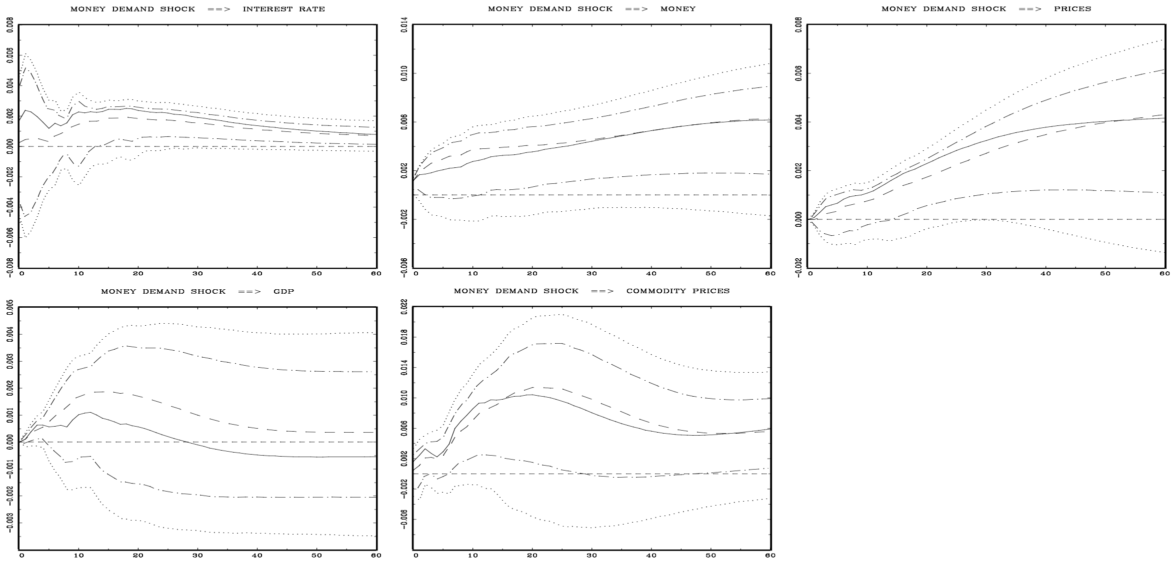

Figure 3 B: LP normalization

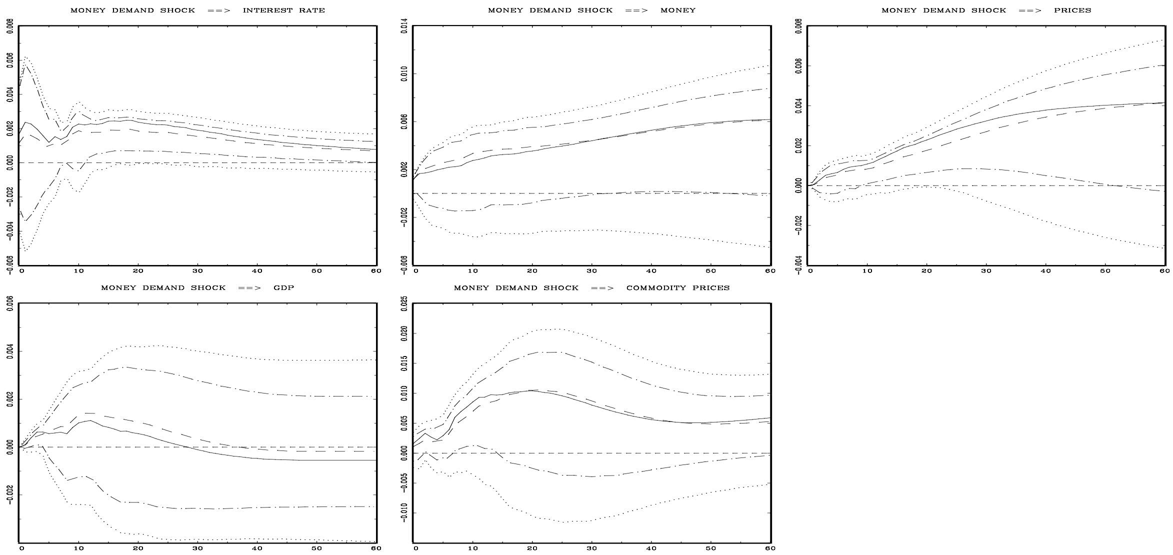

Figure 3 C: PL normalization