

Piffer, Michele

**Working Paper**

## Assessing identifying restrictions in SVAR models

DIW Discussion Papers, No. 1563

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Piffer, Michele (2016) : Assessing identifying restrictions in SVAR models, DIW Discussion Papers, No. 1563, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/130213>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## Discussion Papers

Deutsches Institut für Wirtschaftsforschung

2016

# Assessing Identifying Restrictions in SVAR Models

Michele Piffer

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

#### IMPRESSUM

© DIW Berlin, 2016

DIW Berlin  
German Institute for Economic Research  
Mohrenstr. 58  
10117 Berlin

Tel. +49 (30) 897 89-0  
Fax +49 (30) 897 89-200  
<http://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:  
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:  
<http://ideas.repec.org/s/diw/diwwpp.html>  
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

# Assessing identifying restrictions in SVAR models

Michele Piffer\*

First draft: July 15, 2015

This draft: March 22, 2016

## Abstract

This paper proposes a Bayesian approach to assess if the data support candidate set-identifying restrictions for Vector Autoregressive models. The researcher is uncertain about the validity of some sign restrictions that she is contemplating to use. She therefore expresses her uncertainty with a prior distribution that covers the parameter space both where the restrictions are satisfied and where they are not satisfied. I show that the data determine whether the probability mass in favour of the restrictions increases or not from prior to posterior. Using two applications, I find support for the restrictions used by Baumeister & Hamilton (2015a) in their two-equation model of labor demand and supply, and I find support for the true data generating process in a simulation exercise on the New Keynesian model.

*JEL Classification:* C32, C11.

*Keywords:* Identification, Bayesian Econometrics, Sign Restrictions.

---

\*DIW Berlin, Mohrenstrasse 58, 10117 Berlin, Germany. Email: m.b.piffer@gmail.com, personal web page: <https://sites.google.com/site/michelepiffereconomics/>. I thank Emanuele Bacchiocchi (discussant) and the participants of the summer 2015 CIdE Workshop in Econometrics in Perugia for their encouraging comments on the first draft of this work. I also thank Fabio Canova, Wouter Den Haan, Raffaella Giacomini, Michele Lenza, Helmut Lütkepohl, Haroon Mumtaz, Stefano Neri, Alessia Paccagnini, Max Podstawski and Malte Rieth for further comments and suggestions.

# 1 Introduction

Many research questions in Economics are addressed using Structural Vector Autoregressive models (SVARs). Yet the data only lead the researcher toward a reduced form representation, and give a set of covariance restrictions to map the reduced form model into the structural model. Since the covariance restrictions admit many structural representations, researchers typically complement covariance restrictions with identifying restrictions. For example, one can set-identify demand and supply shocks using the generally accepted view that demand functions are negatively sloped and supply functions are positively sloped.

This paper develops a Bayesian approach to assess whether the data support candidate set-identifying restrictions that the researcher is contemplating to use. The restrictions considered are sign restrictions, which are assessed as follows. Divide the structural parameter space  $\mathcal{A}$  into two subsets,  $\mathcal{A}^{valid}$  and  $\mathcal{A}^{invalid}$ , depending on whether the parameter values satisfy the identifying restrictions or not. For example, in the case of demand and supply functions,  $\mathcal{A}^{invalid}$  includes all structural representations featuring two positively sloped or two negatively sloped equations. The conventional approach is to dismiss  $\mathcal{A}^{invalid}$ , say, by attaching zero prior mass to it. On the contrary, I develop the following preliminary step to the identification of the model, in order to let the researcher assess the support that the data give to the candidate identifying restrictions. The researcher expresses her uncertainty regarding the validity of the restrictions using a prior distribution that attaches mass to both  $\mathcal{A}^{valid}$  and  $\mathcal{A}^{invalid}$ . She then compares the probability mass on  $\mathcal{A}^{valid}$  from prior to posterior. The more this probability mass increases, the more the researcher can subjectively resolve her uncertainty in favour of the identifying restrictions and proceed to structural analysis, for example following Baumeister & Hamilton (2015a,b).

The key challenge faced in developing the above assessment is that the likelihood function is equally high across all candidate structural representations of the VAR. Hence the data support representations that satisfy the identifying restrictions just as much as they support representations that do not satisfy the restrictions. Under a frequentist approach, this implies that, as long as neither  $\mathcal{A}^{valid}$  nor  $\mathcal{A}^{invalid}$  is empty, no assessment of the validity of the restrictions can be developed, unless using additional information external to the model (as for example in Lütkepohl & Netšunajev 2014). On the contrary, the Bayesian approach offers a framework to assess the (subjective) validity attached to the identifying restrictions by studying the probability mass corresponding to  $\mathcal{A}^{valid}$ . This is equivalent to comparing prior and posterior odds associated with  $\mathcal{A}^{valid}$ , relative to the full parameter space  $\mathcal{A}$ .

The use of Bayesian model comparison through prior and posterior odds is well established in the literature. However, to the best of my knowledge, it is not applied to SVAR models, because prior information regarding the relative validity of two specific candidate structures (say the recursive structure and any rotation of it) is not updated by the data, due to the indeterminacy of the model. This point is uncontroversial and is well recognized in the literature since at least the contributions by Kadane (1974), Drèze (1974), Drèze & Richard (1983) and Poirier (1998). However, I show that it holds only for point-identification, not for set-identification, which was introduced to the literature at a later stage (Faust & Leeper 1997, Canova & De Nicoló 2002, Uhlig 2005). By discussing this possibility, the paper aims to fill a gap in the literature, considered the widespread use of set-identified SVAR models.

The reason why the data allow for an update of the prior mass on candidate set-identifying restrictions is that the likelihood function is not flat in the structural parameter space. It only features infinite modes – the infinite candidate structural representations – whose distribution in the structural parameter space reflects the

data covariance structure. Since the latter is identified, the data inform the researcher as to where the candidate representations lay in the parameter space. The model remains unidentified in the sense discussed by Kadane 1974, and no attempt is made to generate a well-behaved posterior by dominating the likelihood with a prior that favours one specific representation. The update occurs because the parameters of the model are not variation-free, due to the constraints that the reduced form parameters impose upon the structural ones. Accordingly, the marginal prior on unidentified parameters can be updated by the data even without prior dependence. This point is already discussed in the literature, but to the best of my knowledge it has not been acknowledged as a tool for model comparison of SVAR models (see for example Koop & Poirier 1997 for an application to switching regression models, and Canova & Sala 2009 and Koop et al. 2013 for an application to DSGE models).

The intuition of the analysis is built using the example of a demand and a supply function for the labour market. The thought experiment is the following. A researcher has data on employment and wages, and is willing to use the identifying restriction that the labor supply function is positively sloped. However, she is uncertain whether she should impose a negatively sloped labor demand function, because economic theory also provides support for positively sloped demands (see for example the efficiency wage literature, Shapiro & Stiglitz 1984 and Yellen 1984). The researcher aims to develop a preliminary assessment of the proposition that, conditioning on one of the equations of the model being positively sloped, the other equation is negatively sloped. If in fact such candidate identifying restriction was not satisfied by the (unknown) process that generated the data, sign restrictions would be inappropriate to disentangle between demand and supply shocks.

I first show that, while being unidentified, the unrestricted model still favours certain parts of the parameter space more than others, in a way summarized in the

covariance structure of the reduced form model. In a bivariate set-up, the feature that drives the update, namely the fact that structural and reduced form parameters are not variation-free, can be isolated analytically through an appropriate reparametrization of the model, as shown. I then apply the analysis to the identification of labor demand and labour supply shocks by Baumeister & Hamilton (2015*a*), and to a standard New Keynesian model. I find extensive support for the identifying restrictions used by Baumeister & Hamilton (2015*a*), and I find that the update successfully guides the researcher in the identification of the New Keynesian model.

Section 2 outlines the proposed Bayesian assessment of set-identifying restrictions in a general framework. Section 3 develops the intuition analytically using the bivariate VAR model for the labour market. Section 4 shows the application to the New Keynesian model. Section 5 concludes.

## 2 Assessing candidate set-identifying restrictions

The structural VAR model can be written as

$$\begin{aligned} Ay_t &= C_1 y_{t-1} + \dots + C_p y_{t-p} + \epsilon_t, \\ &= Cx_t + \epsilon_t. \end{aligned} \tag{1}$$

Matrix  $A$  represents a  $k \times k$  matrix capturing the contemporaneous relations among variables,  $y_t$  is a  $k \times 1$  vector of endogenous variables,  $C = [C_1, \dots, C_p]$  is a  $k \times kp$  matrix that includes the autoregressive parameters of the model,  $x_t = (y'_{t-1}, \dots, y'_{t-p})'$  is a  $kp \times 1$  vector containing the  $p$  lags of the endogenous variables and  $\epsilon_t$  is a  $k \times 1$  vector of serially uncorrelated structural shocks, with  $\epsilon_t \sim N(0, I)$ . The reduced



form model corresponding to (1) is

$$\begin{aligned} y_t &= \Pi_1 y_{t-1} + \dots + \Pi_p y_{t-p} + \nu_t, \\ &= \Pi x_t + \nu_t. \end{aligned} \tag{2}$$

Matrix  $\Pi = A^{-1}C$  contains the reduced form autoregressive parameters of the model, while  $\nu_t = A^{-1}\epsilon_t$  are the reduced form shocks, or VAR innovations, with  $\nu_t \sim N(0, \Omega)$ ,  $\Omega = A^{-1}A'^{-1}$ . Deterministic terms in models (1) and (2) are omitted to simplify notation.

Call  $Y = [y_{-p+1}, \dots, y_0, y_1, \dots, y_T]$  a matrix containing the data. As shown for example in Amisano & Giannini (2002), conditioning on the first  $p$  observations, the likelihood function of the model can be written as

$$L(\Pi, \Omega; Y) = (2\pi)^{-\frac{Tk}{2}} |\Omega|^{-\frac{T}{2}} e^{-\frac{1}{2} \sum_{t=1}^T (y_t - \Pi x_t)' \Omega^{-1} (y_t - \Pi x_t)}. \tag{3}$$

To focus on identification, assume that  $\Pi$  is known (alternatively, suppose we have an estimator for it, say the OLS estimator). The likelihood function can then be written as

$$L(\Omega; \hat{\Omega}) = (2\pi)^{-\frac{Tk}{2}} |\Omega|^{-\frac{T}{2}} e^{-\frac{T}{2} \text{trace}(\Omega^{-1} \hat{\Omega})}, \tag{4}$$

where the data now enter the likelihood through  $\hat{\Omega} = \frac{\sum_{t=1}^T (y_t - \Pi x_t)' (y_t - \Pi x_t)}{T}$ .  $\hat{\Omega}$  corresponds to the OLS/MLE estimator of the unknown matrix  $\Omega$ , and is a consistent estimator for  $\Omega$ . Last, to make the unidentification of the structural model explicit, substitute out the unknown matrix  $\Omega$  with  $A^{-1}A'^{-1}$ , obtaining

$$L(A; \hat{\Omega}) = (2\pi)^{-\frac{Tk}{2}} |A|^T e^{-\frac{T}{2} \text{trace}(A' A \hat{\Omega})}. \tag{5}$$

The unidentification of the structural model (1) can be seen from equation (5). Define  $A_c$  the lower triangular decomposition of  $\hat{\Omega}$ , which implies  $A_c^{-1}A_c'^{-1} = \hat{\Omega}$ . The maximum value of the likelihood function can be reached for  $A = A_c$  from equation (5), or equivalently for  $\Omega = \hat{\Omega}$  from equation (4). Yet, the same maximum is reached for  $A = QA_c$  with  $Q$  any orthogonal matrix, i.e.  $Q'Q = I$ , as can be seen by trivial substitutions into equation (5).  $\Omega$  is identified, since the likelihood has a unique mode in  $\Omega$  (Hamilton 1994). However,  $A$  is not identified, since the likelihood has infinite modes in  $A$ , corresponding to the infinite possible draws of the orthogonal matrix  $Q$ . In this sense, any structural representation of the model is just as consistent with the data, as long as the corresponding matrix  $A$  satisfies the covariance restrictions

$$A^{-1}A'^{-1} = \hat{\Omega}. \quad (6)$$

The literature already discusses the fact that prior information regarding the validity of any specific structural representations is not updated by the data (see Drèze & Richard 1983 for a discussion in the context of Simultaneous Equation Systems). However, while this is true for point-identification, I argue that it is not true for set-identification. To make this point explicit, call  $p(A)$  the prior belief that the researcher expresses regarding the contemporaneous structural parameters of the model.  $p(A)$  can be evaluated anywhere in the parameter space of  $A$ , including in two candidate structural representations of the model, say  $A_c$  and  $QA_c$ . The corresponding prior odds are  $p(A_c)/p(QA_c)$ . Since  $L(A_c; \hat{\Omega}) = L(QA_c; \hat{\Omega})$ , the Bayes Factor  $L(A_c; \hat{\Omega})/L(QA_c; \hat{\Omega})$  equals unity, leading to posterior odds equal to prior

odds,

$$\begin{aligned}\frac{p(A_c|\hat{\Omega})}{p(QA_c|\hat{\Omega})} &= \frac{p(A_c)}{p(QA_c)} \cdot \frac{L(A_c;\hat{\Omega})}{L(QA_c;\hat{\Omega})}, \\ &= \frac{p(A_c)}{p(QA_c)}.\end{aligned}\tag{7}$$

This inability to meaningfully compare candidate specific structural representations is uncontroversial. However, the fact that a unique mode in  $A$  cannot be established does not imply that the data equally support  $A$  anywhere in the parameter space. In fact, the likelihood function is not flat along  $A$ , it only features infinite modes, whose distribution in the parameter space of  $A$  is pinned down by the covariance restrictions (6). Matrices  $A$  achieving the maximum likelihood can be generated without loss of generality starting from  $A_c$  and computing  $A = QA_c$ .  $A_c$  is identified, given that  $\Omega$  is identified and that the decomposition of  $\Omega$  into  $A_c$  is unique (under standard sign normalizations of the shocks). Once she identifies  $A_c$ , the researcher indirectly learns where in the parameter space  $A$  can be, given that  $Q$  cannot be any matrix but it is constrained to be orthogonal.

While the above point is indirectly exploited in several applications, for example in the interpretation of acceptance ratios within frequentist applications of sign restrictions (see Section 3), to the best of my knowledge it is not outlined explicitly as the point of departure for a Bayesian assessment of set-identifying restrictions. To appreciate the relevance of this point, consider again the prior distribution  $p(A)$ . Call  $\mathcal{A}$  the parameter space for  $A$ . Suppose the researcher wants to assess the data support for  $\mathcal{A}^{valid}$ , for example because  $\mathcal{A}^{valid}$  implies impulse responses satisfying certain sign, magnitude or shape restrictions. To avoid trivial situations, consider the case in which there is at least one candidate structural representation in both  $\mathcal{A}^{valid}$  and  $\mathcal{A}^{invalid}$ , i.e.,  $\exists Q$  s.t.  $QA_c \in \mathcal{A}^{valid}$  and  $\exists Q$  s.t.  $QA_c \in \mathcal{A}^{invalid}$ . The researcher

expresses her nondogmatic prior belief favouring  $\mathcal{A}^{valid}$  with  $p(A \in \mathcal{A}^{valid}) < 1$ . Contrary to point identification, the data can potentially speak in favour of or against  $\mathcal{A}^{valid}$ , because the posterior odds,

$$\begin{aligned} \frac{p(A \in \mathcal{A}^{valid} | \hat{\Omega})}{p(A \in \mathcal{A}^{invalid} | \hat{\Omega})} &= \frac{\int_{\mathcal{A}^{valid}} p(A | \hat{\Omega}) dA}{\int_{\mathcal{A}^{invalid}} p(A | \hat{\Omega}) dA}, \\ &= \frac{\int_{\mathcal{A}^{valid}} p(A) L(A, \hat{\Omega}) dA}{\int_{\mathcal{A}^{invalid}} p(A) L(A, \hat{\Omega}) dA}, \end{aligned} \quad (8)$$

do not necessarily equal prior odds  $\frac{p(A \in \mathcal{A}^{valid})}{p(A \in \mathcal{A}^{invalid})}$ , since  $L(A; \hat{\Omega})$  is not flat in  $A$ . Indirectly, the data speak in favour of or against  $\mathcal{A}^{valid}$  by revealing the mass under the likelihood in  $\mathcal{A}^{valid}$ , which is converted into a posterior proper subjective probability through Bayes' rule, as long as the prior distribution  $P(A)$  is proper. Identification uncertainty and estimation uncertainty are then conveniently summarized by the posterior distribution. Note that as a special case, if  $\mathcal{A}^{valid} \equiv A_c$  and  $\mathcal{A}^{invalid} \equiv Q A_c$ , posterior and prior odds coincide, leading the researcher back to the case of point identification from equation (7).

The above discussion outlines the possibility of updating beliefs on candidate set-identifying restrictions by studying the position of the maxima of likelihood function in the parameter space. While the data enter the likelihood function through matrix  $\hat{\Omega}$ , analysing how  $\hat{\Omega}$  can be used to assess the relative support for different parts of the parameter space can be challenging, since covariance restrictions are nonlinear in  $A$ . In the next section, I use a simple bivariate model to show a case in which such assessment can be developed analytically. I use this simplified analysis, which builds on the work by Baumeister & Hamilton (2015a), to relate the proposed approach to the existing literature, and to build the intuition of how the unidentification of the full model does not rule out the possibility of identifying at least some relevant

features of the unknown structure of the data. Section 4 then shows a numerical application to a simulation exercise.

### 3 Application to Baumeister & Hamilton (2015a)

Baumeister & Hamilton (2015a) argue that sign restrictions are best applied using a Bayesian framework that explicitly acknowledges the prior distributions used on the structural parameters of the model. They develop the intuition of their analysis using a bivariate structural VAR model of labour demand and supply. The model is written as

$$\Delta n_t = \alpha \Delta w_t + c_1 + C_{11}(L) \Delta n_{t-1} + C_{12}(L) \Delta w_{t-1} + \epsilon_t^1, \quad (9)$$

$$\Delta n_t = \beta \Delta w_t + c_2 + C_{21}(L) \Delta n_{t-1} + C_{22}(L) \Delta w_{t-1} + \epsilon_t^2. \quad (10)$$

Rewrite the model in compact form as

$$A y_t = c + C_1 y_{t-1} + \dots + C_p y_{t-p} + \epsilon_t, \quad (11)$$

with  $y_t = (\Delta n_t, \Delta w_t)'$  the vector of the growth rate of US total employment and of the real compensation per hour,<sup>1</sup> and

$$A = \begin{pmatrix} -\beta & 1 \\ -\alpha & 1 \end{pmatrix} ; \quad V(\epsilon_t) = D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}.$$

Contrary to model (1), model (11) normalizes one column of matrix  $A$  rather than the covariance matrix of the structural shocks.<sup>2</sup> Baumeister & Hamilton (2015a) identify

---

<sup>1</sup>See footnotes 14 and 15 in the original paper for the description of the data.

<sup>2</sup>In the application in Section 4 I use a model that specifies the contemporaneous impact effect

$\epsilon_t^1$  and  $\epsilon_t^2$  using a Student  $t$  prior distribution for  $\alpha$  truncated at zero and with 95% mass in the support  $[0, 0.22]$ , and a Student  $t$  prior distribution for  $\beta$  truncated at zero and with 95% mass in the support  $[-2.2, 0]$ . Given these sign and magnitude restrictions, they interpret  $\epsilon_t^1$  as a supply shock and  $\epsilon_t^2$  as a demand shock.

To build the intuition behind the assessment of set-identifying restrictions proposed in this paper, I depart from their analysis and develop the following illustrative thought experiment. Suppose the researcher trusts the identifying restriction that the wage elasticity of labour supply is positive, i.e.  $\alpha > 0$ . However, she is uncertain regarding the sign of the elasticity of labour demand  $\beta$ , because economic theory also lends support to positively rather than (more conventional) negatively sloped labour demand functions. This is the case, for example, in the efficiency wage literature, which the literature uses to explain phenomena like involuntary unemployment and dual labour markets (Shapiro & Stiglitz 1984, Yellen 1984).<sup>3</sup> The researcher legitimately expresses uncertainty on whether she should impose the restriction that  $\beta < 0$ , and aims to assess whether the data support such candidate restriction.

More precisely, suppose for simplicity that the researcher is willing to exclude of shocks rather than the contemporaneous relations among variables, i.e.

$$y_t = A^{-1}c + A^{-1}C_1y_{t-1} + \dots + A^{-1}C_p y_{t-p} + A^{-1}\epsilon_t.$$

I also use the normalization on the covariance matrix of the structural shocks. In the application in this section, the normalization of the elements in  $A$  rather than in  $D$  reduces from four to two the dimensionality of the parameters that determine the validity of candidate identifying restrictions, hence allowing for the use of graphs to develop the intuition behind the analysis.

<sup>3</sup>A simplified exposition of this theory is the following. A firm has access to a production function  $y = \tilde{l}^\phi$  (with  $0 < \phi < 1$ ). This production function is standard except for the fact that  $\tilde{l} = e(w)l = cw^\eta l$ , with  $\tilde{l}$  the labour inputs in terms of efficiency units,  $l$  the effective labour input and  $\eta$  the elasticity of labour efficiency with respect to wages. The production function formalizes the fact that workers' productivity is increasing in the real wage  $w$ . In this example, the firm is price taker on the wage market, but anticipates the incentive that the level of real wages exerts on workers'. By solving for labour demand we obtain the wage elasticity of labour demand  $\epsilon_{l^d, w} = -\frac{1-\eta\phi}{1-\phi}$ . If  $\eta > 1/\phi$ , the labour demand is *positively* sloped, because the elasticity of labour efficiency with respect to the real wage is high enough to compensate the firm for the decreasing returns to efficiency units of labour. Alternatively, one can consider negatively rather than positively sloped labor supply functions, which emerge in a model in which the income effect dominates the substitution effect.

values considered extreme for  $\alpha$  and  $\beta$  by using the following support for  $A$ :

$$S_\alpha : 0 \leq \alpha \leq \alpha_H, \quad (12a)$$

$$S_\beta : \beta_L \leq \beta \leq \beta_H, \quad (12b)$$

The boundaries  $\alpha_H, \beta_L$  and  $\beta_H$ , whose relevance in the analysis is discussed below, can be set using either economic theory (as at the end of this section) or reduced form statistics (as in Section 4). The researcher aims to assess the data support for the proposition that, conditioning on  $0 \leq \alpha \leq \alpha_H$ ,  $A$  belongs to  $\mathcal{A}^{valid}$ , with  $\mathcal{A}^{valid} = \{A : 0 \leq \alpha \leq \alpha_H, \beta_L \leq \beta \leq \beta_H\}$ . For the moment, the actual prior distribution  $P(A)$  is not specified other than for its support, nor we introduce  $p(D)$ , the prior on the variance of the structural shocks. This is done in order to emphasize the role of the likelihood function in driving the update.

In Section 2, it was discussed that, while not allowing the researcher to identify a unique matrix  $A$ , the information in  $\hat{\Omega}$  potentially allows the researcher to update her subjective belief on  $p(A \in \mathcal{A}^{valid})$ . In this bivariate model, the main driver of the update can be computed analytically, as is now discussed.

Call  $\hat{\omega}_{ij}$  the  $i, j$  element of  $\hat{\Omega}$ . Define  $\rho$  as

$$\rho(\hat{\Omega}) = \frac{\hat{\omega}_{12}}{\hat{\omega}_{11}} \equiv \frac{\hat{\omega}_{12}}{\sqrt{\hat{\omega}_{11}\hat{\omega}_{22}}} \frac{\sqrt{\hat{\omega}_{22}}}{\sqrt{\hat{\omega}_{11}}}. \quad (13)$$

The covariance restrictions,

$$\hat{\Omega} = A^{-1}DA'^{-1}, \quad (14)$$

impose restrictions on  $\alpha, \beta, d_1, d_2$  given  $\hat{\omega}_{11}, \hat{\omega}_{22}, \hat{\omega}_{12}$ . As shown in the appendix, the system (14) implies that, given  $\hat{\Omega}$ , no solution exists in which  $\alpha$  and  $\beta$  are on the

same side of  $\rho(\hat{\Omega})$ , i.e.

$$\text{if } \alpha > \rho(\hat{\Omega}) \rightarrow \beta < \rho(\hat{\Omega}), \quad (15a)$$

$$\text{if } \alpha < \rho(\hat{\Omega}) \rightarrow \beta > \rho(\hat{\Omega}). \quad (15b)$$

Not all combinations  $\alpha, \beta$  that satisfy conditions (15) are candidate structural representations of the data, they only are if there are values of the remaining structural parameters  $d_1$  and  $d_2$  such that the covariance restrictions in (14) are met. However, no combination of  $\alpha$  and  $\beta$  that fails to satisfy (15) can satisfy (14). Note that  $\rho(\hat{\Omega})$  bears an economic interpretation, as it equals the correlation in the VAR innovations, scaled by a positive term.  $\rho(\hat{\Omega})$  is identified, since it is a function of the elements in  $\Omega$ , which is identified. Asymptotically,  $\rho(\hat{\Omega})$  converges to its population value, as long as the support of  $\rho$  corresponding to  $p(A, D)$  is wide enough to include such value. In finite sample, estimation uncertainty on  $\rho$  is inherited from estimation uncertainty on  $\Omega$ .<sup>4</sup>

Conditions (15) show that the reduced form parameters  $\Omega$  and the structural parameters  $\alpha, \beta$  are not variation-free. This establishes the relation between reduced form and structural parameters that allows for the update of the marginal distribution of the latter given the update of the marginal distribution of the former.<sup>5</sup>

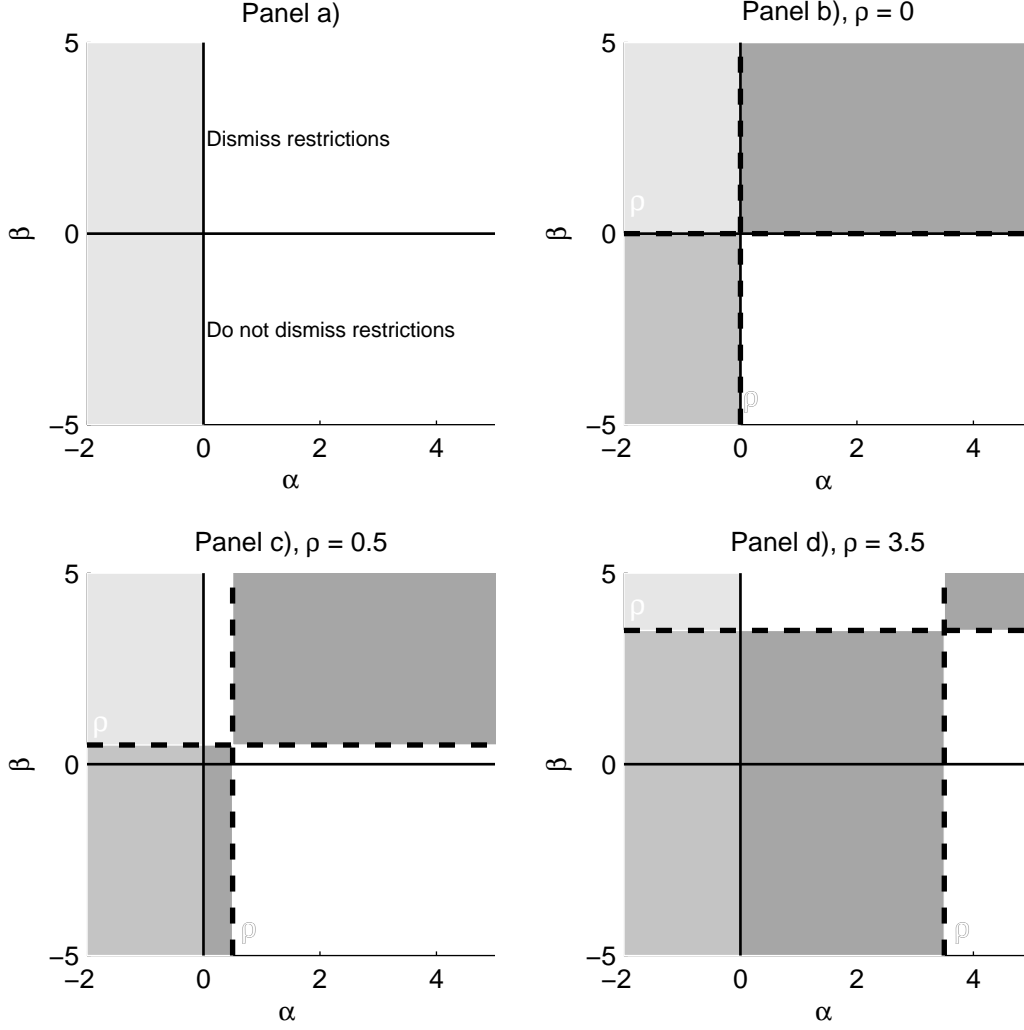
---

<sup>4</sup>Baumeister & Hamilton (2015a) discuss how the sign of the correlation between residuals (hence also the sign of  $\rho(\Omega)$ ) allows interpreting sign restrictions on  $\alpha$  and  $\beta$  as implicit restrictions on the magnitude of the labour elasticities. In this paper I use their framework for model comparison across candidate identifying restrictions, rather than to express identifying restrictions through nondogmatic prior distributions. A related point is made by Paustian (2007), who discusses the fact that if  $\alpha > 0$  is correctly imposed, then the frequentist researcher would always infer that  $\beta < 0$  if  $d_2$  is sufficiently bigger than  $d_1$ , making the correlation (and hence)  $\rho(\Omega)$  negative. I further discuss the relationship of the analysis with the work by Paustian at the end of this section.

<sup>5</sup>As discussed for example in Poirier (1998), consider two parameters,  $\iota$  and  $\nu$ , with prior probabilities  $p(\iota)$  and  $p(\nu)$ . Suppose only  $\iota$  is identified, i.e. the likelihood function is  $L(\iota, \nu; y) = L(\iota; y)$ . The data potentially allow updating the belief on  $\iota$ , i.e.  $p(\iota|y) \neq p(\iota)$ . In addition, it holds that  $p(\nu|\iota, y) = p(\nu|\iota)$ , i.e. for given value of the identified parameter, beliefs about the unidentified parameters cannot be updated by the data. However, the marginal belief on the unidentified parameter can potentially be updated, i.e.  $p(\nu|y) \neq p(\nu)$ , as long as there is a relationship between  $\iota$



Figure 1: Covariance restrictions and structural parameter space



Notes: In Panel a), the light shaded area indicates the subset of the parameter space ruled out through the sign restriction imposed on  $\alpha$ . In the remaining panels, the dark shaded area indicates the subset of the parameter space that cannot be reached by any structural representation of the model, given the value of  $\rho$  implicit in the covariance structure of the data.

Upon learning about  $\rho(\hat{\Omega})$ , the researcher can readjust the distribution of the mass under her prior  $p(A)$  and revise the implicit support to the associated set-identifying

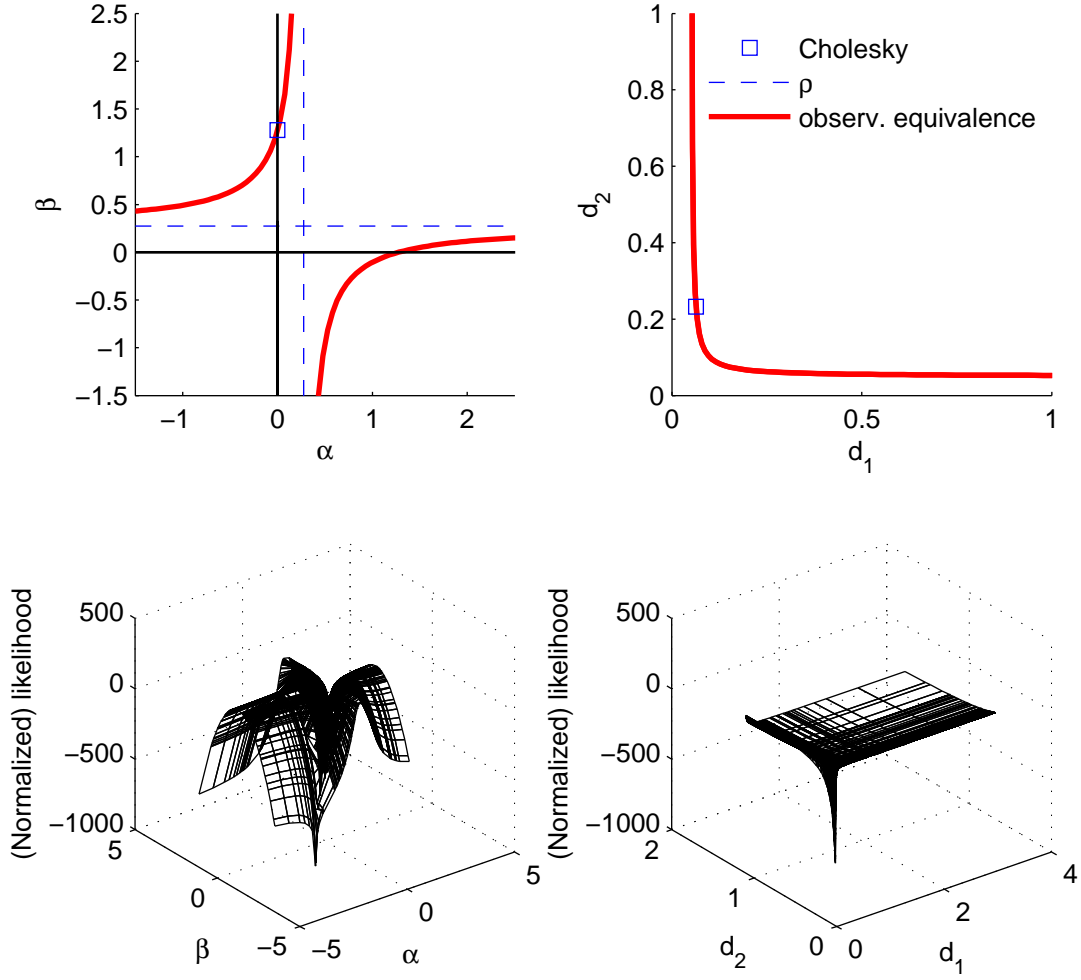
and  $\nu$  that allows the marginal update on  $\iota$  to yield information on  $\nu$ . This occurs either through prior dependence, i.e.  $p(\iota, \nu) \neq p(\iota)p(\nu)$ , or if  $\nu$  and  $\iota$  are not variation-free. An example of the latter is the case in which the support of one parameter depends on the other parameter. In the application to SVARs, the covariance restrictions of the model imply  $A$  and  $\Omega$  being not variation-free. In the bivariate case considered, this constraint on the support of  $\alpha$  and  $\beta$  given  $\Omega$  is highlighted in equations (15).

restrictions. I discuss this graphically in Figure 1. Panel *a* depicts the assessment that the researcher aims to develop, namely the support for  $\beta_L \leq \beta \leq \beta_H$  conditioning on  $0 \leq \alpha \leq \alpha_H$ . The prior distribution  $p(A)$  implies a prior probability favouring the restrictions equal to the mass in correspondence to the bottom right quadrant of panel *a*. The remaining panels show the part of the parameter space that knowledge of  $\hat{\Omega}$  allows to dismiss (dark shaded area), given an implied value of  $\rho$ . If  $\rho(\hat{\Omega}) = 0$ , panel *b*, no candidate representation consistent with  $\hat{\Omega}$  contradicts the candidate identifying restrictions. This implies  $P(A \in \mathcal{A}^{valid}|\hat{\Omega}) > P(A \in \mathcal{A}^{valid})$  for almost any prior distribution  $p(A)$ , since the modes of the likelihood would be in the bottom right quadrant of panel *b*. The case of  $\rho(\hat{\Omega}) = 0$  is the only case in which a frequentist approach is able to deliver an assessment of the validity of the restrictions that  $\beta$  is negative, since  $\mathcal{A}^{invalid}$  would be empty.

Consider now the cases in which  $\rho(\hat{\Omega})$  is positive and small, panel *c*, or relatively large, panel *d*. As shown in the appendix, when  $\rho(\hat{\Omega}) > 0$  one can always find a structural representation in the top right quadrant of panel *a* (even at the cost of  $\alpha$  or  $\beta$  unrealistically diverging to infinity). Accordingly, a frequentist approach would fail to develop an assessment of the candidate identifying restrictions, because any representation in the top right quadrant would be just as likely as any representation in the bottom right quadrant. The information of whether  $\rho(\hat{\Omega})$  is small or big would not be used. On the contrary, under a Bayesian perspective, the information on the magnitude of  $\rho(\hat{\Omega})$  can be used. In fact, the higher is  $\rho(\hat{\Omega})$ , the more the part of the parameter space where the likelihood functions reaches its maxima (non shaded area) overlaps with a region that contradicts the candidate identifying restrictions. Hence, a statistical assessment can be derived by comparing  $P(A \in \mathcal{A}^{valid}|\hat{\Omega})$  and  $P(A \in \mathcal{A}^{valid})$ , or equivalently  $\frac{P(A \in \mathcal{A}^{valid}|\hat{\Omega})}{P(A \in \mathcal{A}^{invalid}|\hat{\Omega})}$  and  $\frac{P(A \in \mathcal{A}^{valid})}{P(A \in \mathcal{A}^{invalid})}$ .

The full set of observational equivalence is shown in Figure 2, which complements

Figure 2: Set of observational equivalence



Notes: The top graphs show the combinations of  $\alpha, \beta, d_1, d_2$  that yield the maximum value of the likelihood function. The lower graphs show the likelihood function in 3D.

Figure 1 (see also the discussion in Baumeister & Hamilton 2015a and Figure 4 in their paper). The top graphs show combinations of  $\alpha, \beta$  (on the left) and  $d_1, d_2$  (on the right) that achieve the maximum likelihood. All combinations along the line are equally likely, as long as the other two parameters are selected from the same candidate structural representation of the model (see the appendix for the computation of this set). The squared symbol in the figure indicates the Cholesky

decomposition  $A_c$ , which then allows for the exploration of the rest of the set of observational equivalence using appropriate rotation and reflection matrices  $Q$  in  $A = QA_c$ . The lower figures display the likelihood function, normalized to 100 at its maximum. The likelihood function has infinite modes, whose distribution along the parameter space reflects the pattern of the covariance structure hidden in  $\hat{\Omega}$ . In the illustrative case of figure 2,  $\rho$  is positive.

So far the prior distribution on  $A$  has been presented only with regard to its support. I now discuss the specification of  $p(A)$  in further detail. Ideally, the prior would faithfully reflect the researcher's uncertainty regarding the validity of the identifying restriction to be assessed, and would leave the Bayesian update to the likelihood function. In practice, under the normalization of  $A$ , the set of observational equivalence goes until  $\pm\infty$  in the parameter space, as clear from Figure 2. This implies that a diffuse prior on  $A$  would leave the marginal likelihood undefined. Similarly, standard distributions like  $t$  Student distributions for  $\alpha$  and  $\beta$  would also be potentially problematic for the purpose of this paper, because they would drive the update toward the parameter space mostly favoured by the prior. On the contrary, the analysis of this paper does not aim to overcome the intrinsic unidentification of the model by means of a proper prior distribution that dominates the likelihood and imposes curvature to the posterior (see the discussion in Kociecki 2013). For this reason, I use independent uniform distributions for  $\alpha, \beta$  and set the boundaries to imply a sufficiently large support for the prior, in a sense discussed below. While in this section the boundaries on the structural parameters rule out part of the set of observational equivalence implied by the data, this does not occur under the normalization on the covariance matrix of the structural shocks used in Section 4, making the use of a uniform distribution less costly.

While being relatively uninformative, even a uniform prior inevitably implies

unintended features, because, as a general note, any uniform distribution of a parameter  $\tau_1$  implies a non-uniform distribution to parameter  $\tau_2 = f(\tau_1)$  whenever  $f(\cdot)$  is a nonlinear function. This nonlinearity indeed emerges in VAR models, where impulse responses and other statistics of interest are nonlinear functions of the parameters of the model. The paper does not aim at deriving less informative priors, but to discuss the ability of the data to still identify selected features of interest regarding the unidentified parameters of a model. See Giacomini & Kitagawa (2015) for a Bayesian robust approach to numerically approximate posteriors under a wide range of prior distributions along a common support.

I conclude this section by applying the analysis to Baumeister & Hamilton (2015*a*). As discussed at the beginning of this section, they use the identifying restrictions that one equation of the model is positively sloped and the other equation negatively sloped. The Bayesian update developed in this paper allows to subjectively assess the validity of the identifying restrictions imposed.

Baumeister & Hamilton (2015*a*) report an estimated matrix  $\hat{\Omega}$  equal to

$$\hat{\Omega} = \begin{pmatrix} 0.5920 & 0.0250 \\ 0.0250 & 0.1014 \end{pmatrix}, \quad (16)$$

for a dataset of  $T = 178$  observations. The corresponding parameter  $\rho$  equals 0.0422, i.e. a positive and small value, as can be inferred from Figure 4 in their paper. The boundaries for the uniform distributions on  $\alpha$  and  $\beta$  are set using economic theory. The truncated prior distribution used by Baumeister & Hamilton (2015*a*) implies 95% prior mass on  $\alpha$  in the support  $[0, 2.2]$ , and 95% prior mass on  $\beta$  in the support  $[-2.2, 0]$ . In the application of this paper I use the support  $[-5, 5]$  for  $\alpha$  and  $\beta$ , which is comparatively a very large support. When conditioning on positive values of  $\alpha$ , the support of  $\alpha$  reduces to  $[0, 5]$ . I use uninformative Gamma priors for  $d_1, d_2$  to avoid

the update from being indirectly driven toward structural representations favoured by the prior distribution on the variances of the structural shocks. The algorithm used to draw from  $p(A|\hat{\Omega})$  is the one proposed by Baumeister & Hamilton (2015a), except that I use an initial grid to ensure that the algorithm explores the entire space of the admissible set of observational equivalence, given the flatness of the posterior coming from a uniform prior. The derivations of the posterior distribution  $p(A|\hat{\Omega})$  are summarized in the appendix, while for the discussion of the Metropolis-Hastings algorithm used the reader is referred to Baumeister & Hamilton (2015a).

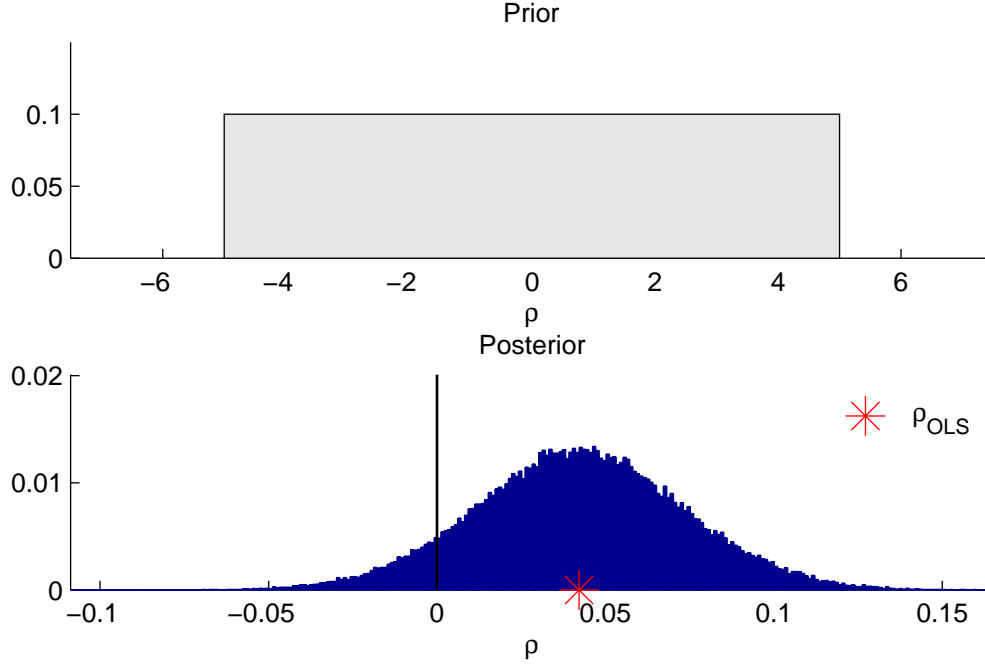
Table 1: Bayesian assessment for the bivariate VAR model

| <i>Hypotheses</i> | <i>Corresponding<br/>parameter space</i>      | <i>Prior<br/>probability</i> | <i>Posterior<br/>probability</i> | <i>Prior<br/>odds</i> | <i>Posterior<br/>odds</i> |
|-------------------|---|------------------------------|----------------------------------|-----------------------|---------------------------|
| <i>I)</i>         | D S $\beta < 0 \mid \alpha > 0$               | 0.5000                       | 0.8650                           | 1.0000                | 6.4069                    |
| <i>II)</i>        | S S $\beta > 0 \mid \alpha > 0$               | 0.5000                       | 0.1350                           | 1.0000                | 0.1560                    |
| <i>III)</i>       | DS $\alpha > 0 \ \& \ \beta < 0$ (or v.versa) | 0.5000                       | 0.8657                           | 1.0000                | 6.4455                    |
| <i>VI)</i>        | DD $\alpha < 0 \ \& \ \beta < 0$              | 0.2500                       | 0.0095                           | 0.3600                | 0.0009                    |
| <i>V)</i>         | SS $\alpha > 0 \ \& \ \beta > 0$              | 0.2500                       | 0.1334                           | 0.3600                | 0.1539                    |

Following the terminology by Zellner (1971), I use the algorithm to compare alternative hypotheses regarding the validity of structural candidate representations. I first study to what extent there is support for a demand shock ( $\beta < 0$ ), given the restriction of the other shock being a supply shock ( $\alpha > 0$ ). This hypothesis, labeled as *Hypothesis I* in Table 1, is compared to the complementary *Hypothesis II* in which the unconditioned shock is another supply shock, i.e. it satisfies the sign pattern associated with supply shock (strict zero equality are omitted for simplicity). *Hypotheses I* versus *II* represent the thought experiment introduced at the beginning of this section to motivate the analysis. I then drop the restriction  $\alpha > 0$  and assess the support for a demand and a supply shock jointly (i.e.  $\alpha > 0$  and  $\beta < 0$  or  $\alpha < 0$  and  $\beta > 0$ ). This hypothesis, labelled as *Hypothesis III*, is compared to the case of

both shocks taking the sign pattern of demand shocks ( $\alpha < 0, \beta < 0$ , *Hypothesis VI*) and to the case of both shocks taking the sign pattern of supply shocks ( $\alpha > 0, \beta > 0$ , *Hypothesis V*).

Figure 3: Identification of  $\rho$  and update of  $p(\rho)$



Notes: The parameter  $\rho$  helps pinning down where the maxima of the likelihood function are in the parameter space.  $\rho$  is uniquely identified, since the likelihood has a single peak in  $\rho$  at its OLS estimate.

Consider *Hypothesis I*. As reported in Table 1, the prior probability mass favouring a demand shock given a supply shock equals 50%, and it is updated upwards to 86%, with prior odds of 1 increasing to posterior odds of 6.40. This suggests that the data provide support for the hypothesis assessed, while *Hypothesis II* is largely dismissed. Consider then *Hypothesis III*, which is the same as *Hypothesis I* except that it does not condition on  $\alpha > 0$ . The prior probability favouring these restrictions is again 50%, which increases to 87%, equivalent to an increase in odds from 1 to 7.12. There is instead limited support for the possibility that  $\alpha$  and  $\beta$  are both

negative or both positive. Support for such parts of the parameter space would be problematic, given the impossibility of sign restrictions to disentangle shocks associated with the same sign pattern of responses. Accordingly, the data seem to lend support to the identifying restrictions used by Baumeister & Hamilton (2015a) to identify the model.

Figure 3 completes the discussion of the results by showing the update of the parameter  $\rho$ . It has been discussed that it is the information recovered by the researcher on this parameter that drives most of the update, since knowledge of this parameter allows the researcher to decrease her subjective probability mass in large parts of the parameter space delimited by  $\rho$ . The prior distributions used for  $\alpha, \beta, d_1, d_2$  imply a prior uniform distribution for  $\rho$  in the support  $[\min(\alpha_L, \beta_L), \max(\alpha_H, \beta_H)]$  (see the appendix), shown in the top graph of Figure 3. The bottom graph shows the posterior. The posterior mode for  $\rho$  equals the value in correspondence to  $\hat{\Omega}$ , i.e. 0.0422. Most importantly, the value is uniquely identified, while the distribution of  $\rho$  reflects estimation uncertainty regarding  $\Omega$ . It is the relatively low but positive value of  $\rho$  that drives the update in favour of *Hypotheses I* and *III* in Table 1.

I conclude this section with a few remarks. As can be inferred from Table 1 and Figure 3, the proposed assessment of candidate set-identifying restrictions only provides the researcher with a probabilistic answer. Except for the special cases, it will not rule out the researcher's uncertainty on the validity of the restrictions, leaving her with the choice of whether to rethink the identification strategy, or to trust the identifying restrictions and proceed with structural analysis. Such degree of arbitrariness could be disciplined by adopting a loss function, and would emerge just as well under a frequentist approach in the choice of a significance level for the test (Zellner 1971, Chapter X).

In some frequentist applications of sign restrictions to VAR models, researchers



report the acceptance ratio, i.e. the ratio of extractions that satisfy the sign restrictions relative to the total number of extractions (see for example Kilian & Murphy 2012 and Straub & Peersman 2006). While what drives the statistical assessment developed in the paper is the same feature that drives the acceptance ratio, there are three important differences between the two approaches. First, in a frequentist setting one can draw inference on the validity of a candidate set of sign restrictions only in the special cases in which the acceptance ratio equals either zero or one. Second, the frequentist use of acceptance ratios does not compare it to the acceptance ratio that is instead implicit in the prior beliefs by the researcher, making it harder to draw conclusions from the results. Third, in its standard use, the acceptance ratio fails to account for estimation uncertainty of the matrix  $\Omega$ . The Bayesian approach, instead, is not constrained to cases in which posterior probabilities equal 0 or 1, it explicitly outlines the prior probability favouring the candidate identifying restrictions, and it naturally combines estimation and identification uncertainty into a proper posterior distribution, ultimately turning the acceptance ratio into a probability.

The difference between frequentist acceptance ratios and Bayesian model comparison allows also to relate the paper to Paustian (2007). Paustian studies to what extent the correct imposition of sign restrictions to a subset of the parameter space implies only representations that meet the correct sign of the unrestricted remaining structural parameters. As in his work, it is the covariance structure of the model that determines whether the candidate structural representations of the model will be in one part or another of the parameter space. The Bayesian analysis pursued in this paper extends his analysis to a probabilistic setting, which allows to draw inference even when draws imply both structural representations with the correct sign of the parameters and structural representations with the incorrect sign of the parameters.

## 4 Application to a New Keynesian model

Having developed the key intuition of the analysis using a bivariate model that allows for analytical derivations of the main driver of the update, I now explore numerically the potential of the Bayesian assessment discussed in Section 2 using a simulation exercise on a standard new Keynesian model.

The model, which is used also in Koop et al. (2013) to study the identification of DSGE models, consists of three equations:

$$R_t = \psi\pi_t + \epsilon_t^M, \quad (17a)$$

$$x_t = E_t(x_{t+1}) - \sigma(R_t - E_t(\pi_{t+1})) + \epsilon_t^D, \quad (17b)$$

$$\pi_t = \beta E_t(\pi_{t+1}) + \gamma x_t + \epsilon_t^S. \quad (17c)$$

Equation (17a) is a monetary policy rule determining the interest rate  $R_t$ . Equation (17b) is an IS curve determining the output gap  $x_t$ . Equation (17c) is a Phillips curve determining inflation  $\pi_t$ . The structural shocks in the model are a monetary shock  $\epsilon^M$ , a demand shock  $\epsilon^D$  and a cost-push shock  $\epsilon^S$ , or supply shock.

Rewriting model (17) in structural VAR form gives

$$R_t = \psi\pi_t + \epsilon_t^M, \quad (18a)$$

$$x_t = -\sigma R_t + \epsilon_t^D, \quad (18b)$$

$$\pi_t = \gamma x_t + \epsilon_t^S, \quad (18c)$$

This, in turn, can be written as

$$Ay_t = \epsilon_t, \quad (19)$$

with

$$A = \begin{pmatrix} 1 & 0 & -\gamma \\ \sigma & 1 & 0 \\ 0 & -\psi & 1 \end{pmatrix} \quad y_t = \begin{pmatrix} R_t \\ x_t \\ \pi_t \end{pmatrix} \quad \epsilon_t = \begin{pmatrix} \epsilon_t^M \\ \epsilon_t^D \\ \epsilon_t^S \end{pmatrix}. \quad (20)$$

I calibrate and simulate the model as in Koop et al. (2013). The values of the structural parameters are  $\sigma = 0.4$ ,  $\gamma = 0.75$ ,  $\psi = 2$ . I then rewrite the model in the specification that formalizes the contemporaneous impact effects of shocks rather than the contemporaneous relationships among variables, since the former is more common in the identification of this type of models. The true structural model generating the data is then

$$y_t = B\epsilon_t, \quad (21)$$

with

$$B = \begin{pmatrix} 0.62 & 0.93 & 0.47 \\ -0.25 & 0.62 & -0.18 \\ -0.50 & 1.25 & 0.62 \end{pmatrix} \quad \mathbf{y}_t = \begin{pmatrix} R_t \\ x_t \\ \pi_t \end{pmatrix} \quad \boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_t^M \\ \epsilon_t^D \\ \epsilon_t^S \end{pmatrix}. \quad (22)$$

The first column of  $B$  corresponds to the impulse vector to a monetary shock, the second column to the impulse vector to a demand shock and the third column to the impulse vector to a supply shock. As in Koop et al. (2013), the shocks are all standard normal and independent of one another, and the artificial data are generated for  $T = 100$  observations. The estimated variance-covariance matrix of the VAR innovations is

$$\hat{\Omega} = \begin{pmatrix} \hat{\omega}_{11} & \hat{\omega}_{21} & \hat{\omega}_{31} \\ \hat{\omega}_{21} & \hat{\omega}_{22} & \hat{\omega}_{23} \\ \hat{\omega}_{31} & \hat{\omega}_{32} & \hat{\omega}_{33} \end{pmatrix} = \begin{pmatrix} 1.48 & 0.34 & 1.15 \\ 0.34 & 0.48 & 0.79 \\ 1.15 & 0.79 & 2.20 \end{pmatrix}. \quad (23)$$

Rewrite matrix  $B$  from model (21) in the general form as

$$B = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix}. \quad (24)$$

Since within this type of models shocks are at best identified up to sign convention, I normalize  $\beta_{1,j}$ ,  $j = 1, 2, 3$  to take positive values, which is achieved through the prior distribution discussed below. Accordingly, each impulse vector  $j$  is associated with only two elements whose sign is unrestricted, i.e.  $\beta_{2,j}$  and  $\beta_{3,j}$ . Each of the three impulse vector can take one of the four sign patterns, summarized in Table 2 and shortened with the letters  $M, D, S$  and  $O$ . Overall, there exist twenty unique combinations of four elements (the four sign patterns  $M, D, S$  and  $O$ ) gathered into groups of three (the three columns of matrix  $B$ ) when considering the order of the shocks as irrelevant. The combinations are  $\{M, D, S\}$ ,  $\{M, M, S\}$ ,  $\{M, D, O\}$ ,  $\dots$ . However, sign restrictions could only detect four of such combinations (namely  $\{M, D, S\}$ ,  $\{O, D, S\}$ ,  $\{M, D, O\}$  and  $\{M, O, S\}$ ), corresponding to the cases in which no sign pattern for the impulse responses is ever repeated.

Table 2: Candidate sign patterns for the shocks

| Interpretation: | <i>Monetary shock</i>                  | <i>Demand shock</i>                    | <i>Supply shock</i>                    | <i>Other shock</i>                     |
|-----------------|--|--|--|--|
| Symbol:         | M                                      | D                                      | S                                      | O                                      |
| Sign pattern:   | $\begin{pmatrix} - \\ - \end{pmatrix}$ | $\begin{pmatrix} + \\ + \end{pmatrix}$ | $\begin{pmatrix} - \\ + \end{pmatrix}$ | $\begin{pmatrix} + \\ - \end{pmatrix}$ |

We know that the true pattern of shocks in the process generating the data consists of one monetary shock, one demand shock and one supply shock, which is the combination  $\{M, D, S\}$ . In real world applications this is of course not known to the researcher, who might express uncertainty regarding the validity of such restrictions.

This could happen, for example, because of economic theories that predict alternative sign patterns for the shocks. For this reason, I develop the thought experiment in which the researcher decides to assess the support by the data for these sign restrictions, instead of immediately using them to run structural analysis.

Operationally, the researcher expresses her uncertainty regarding the validity of some candidate set of sign restrictions through a prior distribution on  $B$ . Under the normalization of the model to  $V(\mathbf{s}_t) = I$  used here, each element  $b_{ij}$  of matrix  $B$  can be interpreted as the impact effect on variable  $i$  of a one standard deviation increase in shock  $j$ . A relatively uninformative way of expressing uncertainty on  $B$  is then to set a prior support for  $b_{ij}$  in the interval  $[-4 \cdot \sqrt{\omega_{ii}}, 4 \cdot \sqrt{\omega_{ii}}]$ . This implies that a one standard deviation shock cannot impact variable  $i$  by more than four times as much as the standard deviation of the innovation of the variable itself. Contrary to the case in Section 3, these boundaries impose no constraint conditioning on a value of  $\Omega$  due to the normalization on the covariance matrix of the structural shocks.<sup>6</sup> The entries of  $B$  are independently distributed, and as in Section 3, are uniformly distributed. For entries  $b_{1,j}$ ,  $j = 1, 2, 3$ , the prior distribution is truncated to be positive. The remaining entries imply positive probability mass to all of the twenty combinations of the four candidate sign patterns summarized in Table 2.

The results of the analysis are summarized in Table 3. The alternative hypotheses

---

<sup>6</sup>In the general case of a  $k \times k$  matrix  $B$  and under the normalization of the covariance of the structural shock to the identity matrix, the covariance restrictions  $\hat{\Omega} = BB'$  imply  $k(k+1)/2$  restrictions. The restrictions corresponding to the diagonal elements of  $\hat{\Omega}$  are  $\hat{\omega}_{ii} = b_{i1}^2 + b_{i2}^2 + \dots + b_{ik}^2$ . Since  $\hat{\omega}_{ii}$  is nonnegative, no element on the right hand side can violate condition

$$-\sqrt{\hat{\omega}_{ii}} \leq b_{ij} \leq \sqrt{\hat{\omega}_{ii}} \quad \text{for } j = 1, 2, \dots, k.$$

Hence, conditioning on a value of  $\hat{\Omega}$ , prior distributions for  $b_{ij}$  in the support  $[-\mu\sqrt{\hat{\omega}_{ii}}, \mu\sqrt{\hat{\omega}_{ii}}]$  with  $\mu > 1$  include the entire set of observational equivalence. In a model with covariance restrictions  $\hat{\Omega} = A^{-1}A'^{-1}$  the constraints on  $a_{ij}$  are  $-\sqrt{\hat{\omega}_{jj}^I} \leq a_{ij} \leq \sqrt{\hat{\omega}_{jj}^I}$ , given  $\hat{\omega}_{jj}^I$  the  $j, j$  entry of  $\hat{\Omega}^{-1}$ . When instead the normalization is applied on a row (in the  $B$  model) or a column (in the  $A$  model) rather than on the covariance matrix of the structural shocks, the above constraints for the unrestricted elements in the matrix  $B$  and  $A$  diverge to the entire real line.

compared are the twenty candidate unique combinations listed on the second column of the table. The line corresponding to the true combination of shocks is highlighted in bold. The analysis is divided into two scenarios. Under *Scenario 1*, the researcher addresses explicitly the possibility that sign patterns of shocks could be non-unique, and that the model might feature shocks which could not be separately set-identified using sign restrictions. *Scenario 2*, instead, conditions the analysis to the candidate combinations of shocks that would be unique in terms of sign patterns. For simplicity, the analysis departs from the same prior distribution and then readjusts the probabilities under *Scenario 2* using the rule of conditional probability.<sup>7</sup> The results are reported as in Table 1, comparing the prior and posterior probability mass in favour of the hypothesis considered, and the corresponding prior and posterior odds. Combinations are listed in descending order according to the posterior probability mass in favour of the candidate hypothesis.

Consider *Scenario 1*. The prior probability mass corresponding to the prior distribution on  $B$  is positive for each candidate combination of shocks, and equals between approximately 0.02 and 0.09.<sup>8</sup> From the posterior probability mass we see that the researcher can confidently rule out 11 candidate combinations, which are the ones to which the algorithm attaches posterior mass equal to zero. The true combination of shocks features a probability mass that increases from 9% to 18%. While not being the combination associated with the highest posterior probability, it ranges as third, already indicating a relatively strong support for such sign pattern, relative to

---

<sup>7</sup>An alternative approach would be to specify the prior distributions on  $b_{ij}$  recursively, restricting new shocks not to take the same pattern of the preceeding shocks.

<sup>8</sup>Under the uniform prior distribution used, each of the four candidate sign patterns of shocks has a 0.25 probability of occurring. This implies a  $0.25 \cdot 0.25 \cdot 0.25 = 0.015625$  probability of drawing a combination of shocks featuring the same sign pattern for all three shocks. The probability mass attached to a combination featuring two equal sign patterns then increases to  $3 \cdot 0.015625 = 0.046875$ , given the three possible permutations of a combination of shocks featuring two identical sign patterns. The probability instead increases to  $6 \cdot 0.015625 = 0.09375$ , given the six possible permutations of a combination of shocks all featuring different sign patterns.

competing patterns.

Table 3: Bayesian assessment for the New Keynesian model

| <i>H<sub>p</sub></i> | <i>Shocks</i> | Scenario 1                   |                              |                             |                             | Scenario 2                   |                              |                             |                             |
|----------------------|---------------|------------------------------|------------------------------|-----------------------------|-----------------------------|------------------------------|------------------------------|-----------------------------|-----------------------------|
|                      |               | <i>Prior</i><br><i>prob.</i> | <i>Post.</i><br><i>prob.</i> | <i>Prior</i><br><i>odds</i> | <i>Post.</i><br><i>odds</i> | <i>Prior</i><br><i>prob.</i> | <i>Post.</i><br><i>prob.</i> | <i>Prior</i><br><i>odds</i> | <i>Post.</i><br><i>odds</i> |
| 1                    | DDM           | 0.05                         | 0.25                         | 0.05                        | 0.34                        |                              |                              |                             |                             |
| 2                    | DDS           | 0.05                         | 0.23                         | 0.05                        | 0.30                        |                              |                              |                             |                             |
| 3                    | <b>MDS</b>    | <b>0.09</b>                  | <b>0.18</b>                  | <b>0.10</b>                 | <b>0.22</b>                 | <b>0.25</b>                  | <b>0.57</b>                  | <b>0.33</b>                 | <b>1.30</b>                 |
| 4                    | DDD           | 0.02                         | 0.09                         | 0.02                        | 0.10                        |                              |                              |                             |                             |
| 5                    | ODS           | 0.09                         | 0.08                         | 0.10                        | 0.08                        | 0.25                         | 0.24                         | 0.33                        | 0.31                        |
| 6                    | DDO           | 0.05                         | 0.06                         | 0.05                        | 0.07                        |                              |                              |                             |                             |
| 7                    | MDO           | 0.09                         | 0.06                         | 0.10                        | 0.07                        | 0.25                         | 0.20                         | 0.33                        | 0.24                        |
| 8                    | MMD           | 0.05                         | 0.02                         | 0.05                        | 0.02                        |                              |                              |                             |                             |
| 9                    | SSD           | 0.05                         | 0.02                         | 0.05                        | 0.02                        |                              |                              |                             |                             |
| 10                   | MOS           | 0.09                         | 0                            | 0.10                        | 0                           | 0.25                         | 0                            | 0.33                        | 0                           |
| 11                   | OOS           | 0.05                         | 0                            | 0.05                        | 0                           |                              |                              |                             |                             |
| 12                   | OOD           | 0.05                         | 0                            | 0.05                        | 0                           |                              |                              |                             |                             |
| 13                   | OOM           | 0.05                         | 0                            | 0.05                        | 0                           |                              |                              |                             |                             |
| 14                   | SSO           | 0.05                         | 0                            | 0.05                        | 0                           |                              |                              |                             |                             |
| 15                   | SSM           | 0.05                         | 0                            | 0.05                        | 0                           |                              |                              |                             |                             |
| 16                   | MMO           | 0.05                         | 0                            | 0.05                        | 0                           |                              |                              |                             |                             |
| 17                   | MMS           | 0.05                         | 0                            | 0.05                        | 0                           |                              |                              |                             |                             |
| 18                   | OOO           | 0.02                         | 0                            | 0.02                        | 0                           |                              |                              |                             |                             |
| 19                   | SSS           | 0.02                         | 0                            | 0.02                        | 0                           |                              |                              |                             |                             |
| 20                   | MMM           | 0.02                         | 0                            | 0.02                        | 0                           |                              |                              |                             |                             |

Inspection of the results under *Scenario 1* shows that the combinations that the algorithm favours relative to the true pattern of shocks are combinations that include two shocks associated with the sign pattern of demand shocks. However, the researcher considering to use sign restrictions might find it suspicious to have two demand shocks, and might be willing to impose the prior restriction that shocks with the same pattern are not in place. This is indeed common practice when using sign restrictions. Note that the problem emerges only with two demand shocks, while any combination of two other shocks (i.e. containing *MM*, *SS* or *OO*) can be ruled out by the researcher based on the unrestricted results from the Bayesian update.

The results for such a conditioning thought experiment are shown under *Scenario 2*. Under this case, the probability mass favouring the true sign pattern of shocks increases from 25% to 57%, and is associated with the highest posterior probability.

Overall, judging whether the Bayesian assessment developed in the paper successfully guides the researcher in identifying the correct pattern of shocks is a subjective matter, and should be explicitly done by setting and defending a full loss function rather than heuristically comparing prior and posterior probabilities. This loss function could impose weights to discipline the choice of favouring a candidate set of restrictions when the posterior probability favouring such restrictions is strictly lower than one, or to discipline the decision between a candidate set of restrictions and other restrictions apparently more favoured by the data but at odds with other prior beliefs of the researcher. The purpose of the paper is not to investigate possible loss functions, but to show that the Bayesian update on candidate set-identifying restrictions can take place, and that there is useful information that the researcher gains from such an exercise.

## 5 Conclusions

Structural interpretations of reduced form vector autoregressive models are usually achieved by combining the covariance restrictions given by the data with additional identifying restrictions. This paper discussed the possibility of statistically assessing set-identifying restrictions using the covariance restrictions of the model. The analysis develops an approach that allows the data to speak in favour of or against candidate set-identifying restrictions, before such restrictions are imposed upon the data when actually identifying the model. The Bayesian model comparison is made possible by the fact that the reduced form and the structural form parameters of the



SVAR model are not variation free, as the two sets of parameters are linked through the covariance restrictions. Since the former are identified, the researcher can update her belief on the latter even without prior dependence between the two.

I first developed the intuition using a simple model in which a reparametrization of the model allows isolating analytically an important driver of the update. I then show the functioning of the proposed Bayesian assessment using two applications. The first one is the bivariate model of the labour market by Baumeister & Hamilton (2015a), the second one is a simulated New Keynesian model. I find extensive support for the sign restrictions used in the identification of the labour market, and I find that the algorithm points the researcher towards the true pattern of shocks driving the New Keynesian model.

The paper mainly relates to the contribution by Baumeister & Hamilton (2015a). Contrary to their work, this paper proposes an approach to achieve model comparison across candidate set-identifying restrictions, rather than to model prior identifying restrictions in a Bayesian way. Overall, the analysis suggests potential in the tool developed, although much more research is required in order to learn how the conclusions drawn from the application discussed hold in a more general framework.

## References

- Amisano, G. & Giannini, C. (2002), *Topics in structural VAR econometrics*, Springer-Verlag.
- Baumeister, C. & Hamilton, J. D. (2015a), ‘Sign restrictions, structural vector autoregressions, and useful prior information’, *Econometrica* **83**(5), 1963–1999.
- Baumeister, C. & Hamilton, J. D. (2015b), ‘Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks’.
- Canova, F. & De Nicoló, G. (2002), ‘Monetary disturbances matter for business fluctuations in the g-7’, *Journal of Monetary Economics* **49**(6), 1131–1159.
- Canova, F. & Sala, L. (2009), ‘Back to square one: identification issues in DSGE models’, *Journal of Monetary Economics* **56**(4), 431–449.
- Drèze, J. H. (1974), Bayesian theory of identification in simultaneous equation models, in S. E. Fienberg & A. Zellner, eds, ‘Studies in Bayesian Econometrics and Statistics’, North-Holland Publishing Company.
- Drèze, J. H. & Richard, J.-F. (1983), ‘Bayesian analysis of simultaneous equation systems’, *Handbook of Econometrics* **1**, 517–598.
- Faust, J. & Leeper, E. M. (1997), ‘When do long-run identifying restrictions give reliable results?’, *Journal of Business & Economic Statistics* **15**(3), 345–353.
- Giacomini, R. & Kitagawa, T. (2015), ‘Robust inference about partially identified SVARs’.
- Hamilton, J. D. (1994), *Time series analysis*, Vol. 2, Princeton University Press.

- Kadane, J. H. (1974), The role of identification in Bayesian theory, *in* S. E. Fienberg & A. Zellner, eds, ‘Studies in Bayesian Econometrics and Statistics’, North-Holland Publishing Company.
- Kilian, L. & Murphy, D. P. (2012), ‘Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models’, *Journal of the European Economic Association* **10**(5), 1166–1188.
- Kociecki, A. (2013), ‘Bayesian approach and identification’.
- Koop, G., Pesaran, M. H. & Smith, R. P. (2013), ‘On identification of Bayesian DSGE models’, *Journal of Business & Economic Statistics* **31**(3), 300–314.
- Koop, G. & Poirier, D. J. (1997), ‘Learning about the across-regime correlation in switching regression models’, *Journal of Econometrics* **78**(2), 217–227.
- Lütkepohl, H. (2007), *New introduction to multiple time series analysis*, Springer Science & Business Media.
- Lütkepohl, H. & Netšunajev, A. (2014), ‘Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs’, *Journal of Applied Econometrics* **29**(3), 479–496.
- Paustian, M. (2007), ‘Assessing sign restrictions’, *The BE Journal of Macroeconomics* **7**(1), 1–31.
- Poirier, D. J. (1998), ‘Revising beliefs in nonidentified models’, *Econometric Theory* **14**(04), 483–509.
- Shapiro, C. & Stiglitz, J. E. (1984), ‘Equilibrium unemployment as a worker discipline device’, *The American Economic Review* **74**(3), 433–444.

- Straub, R. & Peersman, G. (2006), *Putting the New Keynesian Model to a Test (EPub)*, number 6-135, International Monetary Fund.
- 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.
- Yellen, J. L. (1984), ‘Efficiency wage models of unemployment’, *The American Economic Review* pp. 200–205.
- Zellner, A. (1971), *An introduction to Bayesian inference in econometrics*, John Wiley & Sons, Inc.

## Appendix A: Decomposition of the covariance restrictions in Section 2

The covariance restrictions of the model in Section 2 are

$$\Omega = A^{-1}DA'^{-1}.$$

Define  $\omega_{i,j}$  the  $i, j$  entry of matrix  $\Omega$  (or of  $\hat{\Omega}$ , if the analysis is in sample instead of in population) and  $d_i$  the  $i, i$  entry of the diagonal matrix  $D$ . Under the normalization to unity of the second column of  $A$  and the parametrization  $A = \begin{pmatrix} -\beta & 1 \\ -\alpha & 1 \end{pmatrix}$ , the covariance restrictions are summarized by the following system of equations:

$$\omega_{11} = \frac{1}{(\alpha - \beta)^2} [d_1 + d_2], \quad (25a)$$

$$\omega_{12} = \frac{1}{(\alpha - \beta)^2} [\alpha d_1 + \beta d_2], \quad (25b)$$

$$\omega_{22} = \frac{1}{(\alpha - \beta)^2} [\alpha^2 d_1 + \beta^2 d_2]. \quad (25c)$$

This is a system of four unknowns  $(\alpha, \beta, d_1, d_2)$  in three elements  $(\omega_{11}, \omega_{12}, \omega_{22})$ . Data (if working in sample) or population moments (if working in population) enter the restrictions through  $(\omega_{11}, \omega_{12}, \omega_{22})$ .

To solve for the structural form, obtain  $d_1 = (\alpha - \beta)(\omega_{12} - \beta\omega_{11})$  as a solution for  $d_1$  in  $\alpha, \beta$  from equations (25a) and (25b). Then obtain  $d_1 = \alpha^{-1}(\alpha - \beta)(\omega_{22} - \beta\omega_{12})$  as a solution for  $d_1$  in  $\alpha, \beta$  from equations (25b) and (25c). Equating these expressions gives the solution for  $\beta$  for an arbitrary value of  $\alpha$ :

$$\beta(\alpha) = \frac{\omega_{22} - \alpha\omega_{12}}{\omega_{12} - \alpha\omega_{11}}. \quad (26)$$

Equation (26) coincides with equation (51) in Baumeister & Hamilton (2015*a*). Each combination of  $\alpha$  and  $\beta(\alpha)$  is associated with a combination of  $d_1$  and  $d_2$  that ensures that the covariance restrictions are met.  $d_1$  and  $d_2$  can be computed as

$$d_1(\alpha, \beta) = (\alpha - \beta)(\omega_{12} - \beta\omega_{11}), \quad (27)$$

$$d_2(\alpha, \beta) = (\alpha - \beta)(\alpha\omega_{11} - \omega_{12}). \quad (28)$$

Equations (26), (27), (28) jointly provide the solution for the structural parameters  $\beta, d_1, d_2$  given an arbitrary choice of  $\alpha$ .

Define

$$\rho = \frac{\omega_{12}}{\omega_{11}}. \quad (29)$$

From equations (27) and (28) (or dividing (25b) by (25a) and rearranging) we obtain

$$\delta(\alpha, \beta) \equiv \frac{d_1(\alpha, \beta)}{d_1(\alpha, \beta) + d_2(\alpha, \beta)} = \frac{\rho - \beta}{\alpha - \beta}. \quad (30)$$

Since the left-hand side is constrained to the interval  $[0, 1]$ ,  $\rho$  constrains  $\alpha$  and  $\beta$ . If  $\alpha > \beta$ , then conditions  $\delta > 0$  and  $\delta < 1$  jointly imply  $\beta < \rho < \alpha$ . If instead  $\alpha < \beta$ , then conditions  $\delta > 0$  and  $\delta < 1$  jointly imply  $\alpha < \rho < \beta$ . Hence, no solution exists in which  $\alpha$  and  $\beta$  are on the same side of  $\rho$ . Equality between  $\alpha$  and  $\beta$  would lead  $\delta(\alpha, \beta)$  to diverge to infinity, or equivalently, to the singularity of matrix  $A$ . Combining (26) with (29) and rearranging gives

$$\beta(\alpha) = \frac{\frac{\omega_{22}}{\omega_{11}} - \alpha\rho}{\rho - \alpha} \quad \text{or} \quad \alpha(\beta) = \frac{\frac{\omega_{22}}{\omega_{11}} - \beta\rho}{\rho - \beta}, \quad (31)$$

which shows that  $\beta$  or  $\alpha$  diverge when  $\alpha$  and  $\beta$  approach  $\rho$ , yielding  $\delta(\alpha, \beta)$  to approach 1 or 0, respectively. A solution can instead always be computed for  $\alpha \neq \rho$

or  $\beta \neq \rho$ . To compute the implicit prior distribution on  $\rho$  given  $p(A, D)$ , combine equations (25a) and (25b) with (29), obtaining

$$\rho = \alpha \frac{d_1}{d_1 + d_2} + \beta \frac{d_2}{d_1 + d_2}. \quad (32)$$

This shows that, given prior distributions  $\alpha \sim U[\alpha_L, \alpha_H]$  and  $\beta \sim [\beta_L, \beta_H]$ , the support for  $p(\rho)$  is  $[\min(\alpha_L, \beta_L), \max(\alpha_H, \beta_H)]$ . The uniform behaviour of  $\rho$  in this support under independent uninformative  $\Gamma$  distributions for  $d_1$  and  $d_2$  is checked with simulations.

## Appendix B: Derivations of the posterior distribution

In the general case in which the covariance matrix  $D$  of the structural shocks is not normalized to the identity matrix, the likelihood function (5) from Section 2 rewrites as<sup>9</sup>

$$L(A, D; \hat{\Omega}) = (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot |D|^{\frac{T}{2}} \cdot e^{-\frac{T}{2} \text{trace}(A\hat{\Omega}A'D^{-1})}. \quad (33)$$

Call  $V(A) = A\hat{\Omega}A'$ , and  $V_i(A)$  the  $i, i$  entry of  $V(A)$ . Since  $D^{-1} = \text{diag}(d_1^{-1}, \dots, d_n^{-1})$  and since the determinant of a diagonal is the product of its diagonal entries, the

---

<sup>9</sup>In Baumeister & Hamilton (2015a) the data do not enter via  $\hat{\Omega}$  but via data  $\{y_t\}_{t=1}$ , as the model also specifies a prior distribution on the autoregressive component of the SVAR. Since under the normal prior distribution on the latter component of the model they show that the marginal posterior distribution for  $A$  depends on the data only via  $\hat{\Omega}$ , it is without loss of generality to develop the assessment in this paper starting from  $\Omega$  rather than from the full specification of the model.

likelihood becomes

$$\begin{aligned} p(\hat{\Omega}|A, D^{-1}) &= (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot \left( \prod_{i=1}^k d_i^{-1} \right)^{\frac{T}{2}} \cdot e^{-\frac{T}{2} \sum_{i=1}^k V_i(A) \cdot d_i^{-1}}, \\ &= (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot \prod_{i=1}^k \left( (d_i^{-1})^{\frac{T}{2}} \cdot e^{-\frac{T}{2} V_i(A) \cdot d_i^{-1}} \right). \end{aligned}$$

Consider the prior distribution

$$p(A, D^{-1}) = p(A) \cdot p(D^{-1}),$$

with some prior  $p(A)$  for  $A$  and  $d_i^{-1}$  independently Gamma distributed with shape and rate parameters  $\psi_{s,i}$  and  $\psi_{r,i}$ , respectively. The pdf of  $d_i^{-1}$  is

$$\begin{aligned} p(d_i^{-1}) &= \frac{1}{\Gamma(\psi_{s,i})} \psi_{r,i}^{\psi_{s,i}} (d_i^{-1})^{\psi_{s,i}-1} e^{-\psi_{r,i} d_i^{-1}}, \\ &\propto (d_i^{-1})^{\psi_{s,i}-1} e^{-\psi_{r,i} d_i^{-1}}. \end{aligned}$$

The joint posterior is then written as

$$\begin{aligned} p(A, D^{-1}|\hat{\Omega}) &= p(A) \cdot \prod_{i=1}^k \left( \frac{1}{\Gamma(\psi_{s,i})} \psi_{r,i}^{\psi_{s,i}} \cdot (d_i^{-1})^{\psi_{s,i}-1} \cdot e^{-\psi_{r,i} d_i^{-1}} \right) \cdot \\ &\quad (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot \prod_{i=1}^k \left( (d_i^{-1})^{\frac{T}{2}} \cdot e^{-\frac{T}{2} V_i(A) \cdot d_i^{-1}} \right) \cdot p(\hat{\Omega})^{-1}, \end{aligned}$$

which can be simplified to

$$\begin{aligned} p(A, D^{-1}|\hat{\Omega}) &= p(A) \cdot (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot p(\hat{\Omega})^{-1} \cdot \\ &\quad \prod_{i=1}^k \left( \frac{1}{\Gamma(\psi_{s,i})} \cdot \psi_{r,i}^{\psi_{s,i}} \cdot (d_i^{-1})^{\psi_{s,i}-1+\frac{T}{2}} \cdot e^{-[\psi_{r,i} + \frac{T}{2} V_i(A)] d_i^{-1}} \right). \end{aligned}$$



After substituting  $\psi_{s,i}^* = \psi_{s,i} + \frac{T}{2}$  and  $\psi_{r,i}^*(A) = \psi_{r,i} + \frac{T}{2} \cdot V_i(A)$  and rearranging, we get

$$p(A, D^{-1}|\hat{\Omega}) = p(A) \cdot (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot p(\hat{\Omega})^{-1} \cdot \prod_{i=1}^k \left( \frac{\Gamma(\psi_{s,i}^*)}{\psi_{r,i}^*(A)^{\psi_{s,i}^*}} \cdot \frac{\psi_{r,i}^{\psi_{s,i}}}{\Gamma(\psi_{s,i})} \right) \cdot \underbrace{\prod_{i=1}^k \left( \frac{1}{\Gamma(\psi_{s,i}^*)} \cdot \psi_{r,i}^*(A)^{\psi_{s,i}^*} \cdot (d_i^{-1})^{\psi_{s,i}^*-1} \cdot e^{-\psi_{r,i}^*(A)d_i^{-1}} \right)}_{p(d_i^{-1}|A, \hat{\Omega})}. \quad (34)$$

This implies

$$d_i^{-1}|A, \hat{\Omega} \sim \Gamma(\psi_{s,i}^*, \psi_{r,i}^*(A)). \quad (35)$$

From this result, since  $p(A, D^{-1}|\hat{\Omega}) = p(A|\hat{\Omega}) \cdot p(D^{-1}|A, \hat{\Omega})$ , the solution for  $p(A|\hat{\Omega})$  coincides with the first part of the equation (34), i.e.

$$\begin{aligned} p(A|\hat{\Omega}) &= p(A) \cdot (2\pi)^{-\frac{Tk}{2}} \cdot |A|^T \cdot p(\hat{\Omega})^{-1} \cdot \prod_{i=1}^k \left( \frac{\Gamma(\psi_{s,i}^*)}{\Gamma(\psi_{s,i})} \cdot \frac{\psi_{r,i}^{\psi_{s,i}}}{\psi_{r,i}^*(A)^{\psi_{s,i}^*}} \right), \\ &= \underbrace{p(\hat{\Omega})^{-1} \cdot (2\pi)^{-\frac{Tk}{2}} \cdot \prod_{i=1}^k \left( \frac{\Gamma(\psi_{s,i}^*)}{\Gamma(\psi_{s,i})} \cdot \psi_{r,i}^{\psi_{s,i}} \right)}_{\kappa} \cdot p(A) \cdot |A|^T \cdot \prod_{i=1}^k \left( \frac{1}{\psi_{r,i}^*(A)^{\psi_{s,i}^*}} \right), \\ &= \kappa \cdot p(A) \cdot |A|^T \cdot \prod_{i=1}^k \left( \frac{1}{\psi_{r,i}^*(A)^{\psi_{s,i}^*}} \right), \\ &= \kappa \cdot p(A) \cdot |A|^T \cdot \prod_{i=1}^k \left( \frac{1}{[\psi_{r,i} + \frac{T}{2} \cdot V_i(A)]^{\psi_{s,i} + \frac{T}{2}}} \right). \end{aligned}$$

Under a non informative prior on  $d_i^{-1}$ ,  $\psi_{s,i} = \psi_{r,i} = 0$ , implying

$$\begin{aligned}
p(A|\hat{\Omega}) &= \kappa \cdot p(A) \cdot |A|^T \cdot \prod_{i=1}^k \left( \frac{1}{[\frac{T}{2} \cdot V_i(A)]^{\frac{T}{2}}} \right), \\
&= \underbrace{\kappa \cdot \frac{2^k}{T^k}}_{\tilde{\kappa}} \cdot p(A) \cdot |A|^T \cdot \prod_{i=1}^k \left( \frac{1}{[V_i(A)]^{\frac{T}{2}}} \right), \\
&= \tilde{\kappa} \cdot p(A) \cdot |A|^T \cdot \frac{1}{|diag(A\hat{\Omega}A')|^{\frac{T}{2}}}, \\
&= \frac{\tilde{\kappa} \cdot p(A) \cdot |A|^T}{|diag(A\hat{\Omega}A')|^{\frac{T}{2}}}.
\end{aligned}$$

where, again, we used the fact that the determinant of a diagonal matrix is the product of its diagonal elements. This equation coincides with equation (22) in Baumeister & Hamilton (2015a), after noting that  $|\hat{\Omega}|$  is not a function of  $A$  and hence is included in the term  $\tilde{\kappa}$ .

For Section 4, the structural VAR model is written specifying the impact effect of the structural shocks rather than the contemporaneous relationships among variables. For this so-called  $B$  model  $y_t = \Pi(L)y_{t-1} + B\epsilon_t$  (Lütkepohl 2007), the likelihood function equals

$$L(B, D; \hat{\Omega}) = (2\pi)^{-\frac{Tk}{2}} \cdot |B|^{-T} \cdot |D|^{\frac{T}{2}} \cdot e^{-\frac{T}{2} trace(B^{-1}\hat{\Omega}B'^{-1}D^{-1})}. \quad (36)$$

Calling  $V(B) = B^{-1}\hat{\Omega}B'^{-1}$ , the same derivations above give

$$d_i^{-1}|B, \hat{\Omega} \sim \Gamma(\psi_{s,i}^*, \psi_{r,i}^*(B)). \quad (37)$$

with  $\psi_{s,i}^* = \psi_{s,i} + \frac{T}{2}$ , and  $\psi_{r,i}^*(B) = \psi_{r,i} + \frac{T}{2} \cdot V_i(B)$  and

$$p(B|\hat{\Omega}) = \kappa \cdot p(B) \cdot |B|^{-T} \cdot \prod_{i=1}^k \left( \frac{1}{[\psi_{r,i} + \frac{T}{2} \cdot V_i(B)]^{\psi_{s,i} + \frac{T}{2}}} \right).$$

Under a noninformative prior for  $d_i^{-1}$ ,  $p(B|\hat{\Omega})$  equals

$$p(A|\hat{\Omega}) = \frac{\tilde{\kappa} \cdot p(B) \cdot |B|^{-T}}{|diag(B^{-1}\hat{\Omega}B'^{-1})|^{\frac{T}{2}}}.$$

Under the normalization  $D = I$  used in Section 3, the element in  $D$  drops from equation (36) and the posterior distribution for  $B$  is derived as

$$p(B|\hat{\Omega}) = p(B)(2\pi)^{-\frac{Tk}{2}} \cdot |B|^{-T} \cdot e^{-\frac{T}{2}trace(B^{-1}\hat{\Omega}B'^{-1})} p(\hat{\Omega})^{-1}. \quad (38)$$

Regarding posterior simulation, the marginal posterior densities  $p(A|\hat{\Omega})$  and  $p(B|\hat{\Omega})$  derived above can be used in a Metropolis-Hastings algorithm to draw posterior draws. To do so, I follow the algorithm by Baumeister & Hamilton (2015a), with one important modification. In Baumeister & Hamilton (2015a), curvature to the posterior is achieved through  $t$  Student priors, which ensure that the posterior is unimodal. However, the use of a uniform prior in this paper implies that the posterior has infinite modes, representing the underidentification of the model. To reduce the risk of the algorithm exploring only part of the posterior, in Section 3 I do not initialize the Metropolis-Hastings algorithm on a mode of the posterior, but rather on 50 random grip points, computed numerically on  $\hat{\Omega}$  to preliminary inspect where the set of observational equivalence lays in the parameter space. I then run as many Metropolis-Hastings algorithms as many grid points considered and store all the extractions, ensuring that the acceptance rate is around 30% for all points considered. A diagnosis of the coverage of the set of observational equivalence is done

by generating 500 random candidate decompositions of  $\hat{\Omega}$  and by ensuring that the Euclidean distance of each posterior extraction from each candidate decomposition are all relatively small and comparable, i.e. that there are no clear outliers.