Assessing identifying restrictions in SVAR models

Michele Piffer*

February 14, 2016

Abstract

This paper proposes a Bayesian approach to assess if the data support candidate set-identifying restrictions in Vector Autoregressive models. I study the case of sign restrictions. The researcher expresses her uncertainty regarding the validity of the restrictions using a prior distribution that covers the parameter space both where the restrictions are satisfied and where they are not satisfied. The correlation in the data then determines whether the probability mass in favour of the restrictions increases or not from prior to posterior. I apply the proposed Bayesian assessment to a two-equation model of labour demand and supply, and to a 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 Fabio Canova, Raffaella Giacomini, Michele Lenza, Helmut Lütkepohl, Stefano Neri, Alessia Paccagnini, Max Podstawski and Malte Rieth for comments and suggestions. I also thank Emanuele Bacchiocchi and the participants to the 3^{rd} CIdE Workshop in Econometrics in Perugia, August 27-28, 2015, and the participants to a seminar at Free University, Berlin.

1 Introduction

Several research questions in Economics are addressed using structural vector autoregressive models (SVARs). Yet the data only lead the researcher towards 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 potentially conflicting 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 and supply functions are negatively and positively sloped, respectively.

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

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 $\mathcal{A}^{rest.\ valid}$ nor $\mathcal{A}^{rest.\ invalid}$ are empty, no assessment of the validity of the candidate identifying 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 possible framework to assess the (subjective) validity attached to the candidate identifying restrictions by studying the probability mass corresponding to $\mathcal{A}^{rest.\ valid}$. This is equivalent to comparing prior and posterior odds associated with $\mathcal{A}^{rest.\ valid}$, 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 that studies identified models. On the contrary, to the best of my knowledge, it has not been 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 undeterminacy of the model. While this point is uncontroversial and has been recognized in the literature at least since the contributions by Dreze (1974) and Poirier (1998), I show that it holds only for point-identification, not for set-identification, as for instance with the popular sign restrictions. By discussing this possibility, the paper aims to fill a gap in the literature, considered the widespread use of set-identified VAR models.

The key intuition of why the data allow for the update of the prior mass on sets of the parameter space, yet not on specific point-identifying candidate restrictions, is that the likelihood function is not completely flat in the parameter space. It only features infinite modes (i.e. infinite candidate structural representations along the maximum of the likelihood) whose distribution in the structural parameter space reflects the data covariance structure, which *is* identified. Hence, while not uniquely identifying a specific structural representation of the model, the data still inform the researcher as to where the candidate representations lay in the parameter space. The model remains by construction unidentified, and no attempt is made to generate a well-behaved posterior by dominating the likelihood with a prior that favours one specific structural representation. The update is mainly driven by the fact that the structural parameters are not variation-free, due to the non-linear constrains that the reduced form parameters impose upon the structural ones (see for example Canova & Sala 2009 and Koop et al. 2013 for a discussion on this point in the analysis of 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 labour and real wage, and is willing to use the identifying restriction that a supply function is positively sloped. However, she is uncertain whether she should impose a negatively sloped demand function, because economic theory also provides support for positively sloped demand functions (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 inconsistent with the (unknown) process that generated the data, sign restrictions would be inappropriate to disentangle demand versus supply shocks.

I first show analytically that, while being unidentified, the unrestricted model still favours certain parts of the parameter space more than others through the information in the covariance structure of the reduced form model. This, in turn, allows the likelihood function to drive the update of any prior distribution along the structural parameter space. I then use two simulation exercises to study how the update of a uniform prior can provide useful information to the researcher regarding her decision to impose the candidate identifying restrictions and proceed to structural analysis. Last, I apply the algorithm to the analysis by Baumeister & Hamilton (2015) and to a standard New Keynesian model. I find extensive support for the identifying restrictions used by Baumeister & Hamilton (2015), and I find that the update successfully guides the researcher in the identification of the New Keynesian model.

The main contribution of the paper is derived in Section 2 using analytical derivations and two simulation exercises. Having outlined the contribution of the paper, I then move to Section 3 with a discussion of the approach proposed and of its relation to the literature. Section 4 shows the application to the New Keynesian model. Section 5 concludes.

2 Assessing candidate set-identifying restrictions

This section outlines the key intuition of the paper, which is that, while the relative validity of specific candidate structural VAR representations is not updated by the data, the update occurs when considering set-identifying candidate restrictions, for example sign restrictions. I build the intuition using a VAR model with two variables.

2.1 A VAR model of labour demand and supply

Consider a standard bivariate structural VAR model, which can be written as

$$A\boldsymbol{y}_t = \boldsymbol{c} + C(L)\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t. \tag{1}$$

In equation (1), $\boldsymbol{y}_t = (l_t, w_t)'$ is a 2×1 vector of the endogenous variables, namely labour and real wage (in logs), $\boldsymbol{c} = (c_1, c_2)'$ is a vector of constants, $C(L)$ is a matrix of lag polinomials capturing the autoregressive component of the model, and $\boldsymbol{\epsilon}_t$ is a vector of structural shocks, with $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, D)$. Bold letters indicate vectors. As in Baumeister & Hamilton (2015), the model is normalized to unity on a column of matrix A, leaving matrix D unrestricted:

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

This normalization helps reducing from four to two the dimensionality of the parameters which pin down the the validity of the restrictions considered. The normalized model has four structural parameters, which are summarized in the vector $\boldsymbol{\delta} = (\alpha, \beta, d_1, d_2)'$. The two structural equations of the model feature an elasticity of labour with respect to real wage equal to α and β. Absent guidance from economic theory, these elasticities cannot be interpreted in a structural sense.

The data restrict the structural identification of the model through the variance-covariance matrix of the reduced form shocks $\boldsymbol{u}_t = A^{-1}\boldsymbol{\epsilon}_t$. Formally,

$$\Sigma = A^{-1}DA'^{-1}, \tag{2}$$

$$\text{with} \;\; \Sigma = V(\boldsymbol{u}_t) = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{pmatrix}.$$

The reduced form parameters $\boldsymbol{c}, C(L), \Sigma$ are uniquely identified (Hamilton 1994). However, given an estimate for Σ, there are infinite sets $\boldsymbol{\delta}_i = (\alpha_i, \beta_i, d_{1,i}, d_{2,i})'$ of structural parameters that are equally consistent with the data, because they all correspond to structural representations that achieve the maximum value of the likelihood function. As an illustration, a standard frequentist way of inspecting candidate representations would be to compute the Cholesky decomposition of an estimate of Σ, apply rotations or similar orthogonal transformations, and impose the normalization on the first column of A.

Suppose the researcher trusts the identifying restriction that the wage elasticity of labour supply is positive. She hence considers imposing the restriction that, say, $\alpha > 0$. However, she is uncertain regarding the sign of the elasticity of labour demand, 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 has been used to explain phenomena like involuntary unemployment and dual labour markets, and which is simplified in the following example (Yellen 1984). 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 η 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.

As the example illustrates, the researcher could legitimately express uncertainty on whether she should impose the restriction that $\beta < 0$, and aims to develop a preliminary probabilistic assessment of the following proposition:

Proposition I

$\beta < 0 \mid \alpha > 0$: *conditioning of α being positive, β is negative*

The discussion is best outlined graphically. In Figure 1 panel a, the researcher wants to assess the support for the representations in the top-right quadrant relative to the representations bottom-right quadrant. If Proposition I is inconsistent with the (unknown) model generating the data, sign restrictions would not be suitable to identify the model. The light shaded area in the figure imposes the sign restrictions on the supply function, which the researcher is willing to impose.

To develop an intuition of how such an assessment can be developed in a Bayesian framework, start from the restrictions that the reduced form parameters $\sigma_{11}^2, \sigma_{12}^2, \sigma_{22}^2$ impose on the structural paramenters α, β, d_1, d_2. The covariance restrictions (2) contain three restrictions.[1] As shown in the appendix, the restrictions imply that no candidate structural representation exists such that α and β are on the same side of $\rho = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}\frac{\sigma_{11}}{\sigma_{22}}$, i.e.

$$\begin{aligned} &\text{if } \alpha \geq \rho \quad \rightarrow \quad \beta < \rho, \\ &\text{if } \alpha \leq \rho \quad \rightarrow \quad \beta > \rho. \end{aligned} \tag{3}$$

Not all combinations α, β that satisfy conditions (3) are candidate structural representations, they only are if there are values of the remaining structural parameters d_1 and d_2 such that the covariance restrictions in (2) are met. However, no combination of α and β that fails to satisfy (3) can satisfy (2). Note that ρ bears an economic interpretation, as it equals the correlation in the VAR innovations, scaled by a positive term.

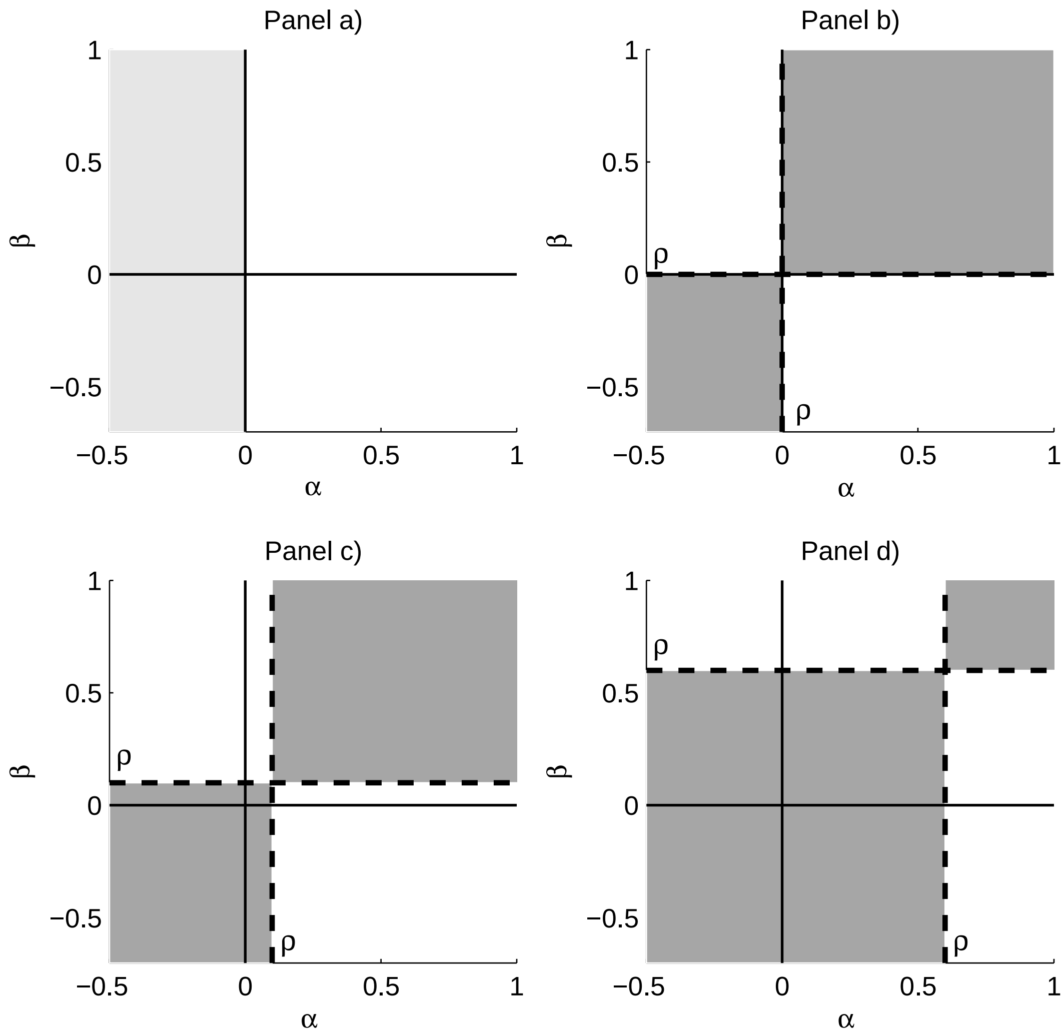
Consider the case of $\rho = 0$, panel *b*. In this case, no candidate representation consistent with the estimated matrix Σ contradicts the candidate identifying restrictions, because there could be no mass in correspondence to the top-right quadrant of panel *a*, while the top-left quadrant would be excluded due to the restriction imposed on α. Under a frequentist perspective, the researcher could indeed conclude in favour of Proposition I and proceed with structural analysis.

Consider now the cases in which ρ is positive and very small, panel *c*, or relatively large, panel *d*. As shown in the appendix, when $\rho > 0$ one can always find a structural

[1] The restrictions are

$$\begin{aligned} \sigma_{11}^2 &= \frac{1}{(\alpha-\beta)^2}[\beta^2 d_1 + \alpha^2 d_2], \\ \sigma_{12}^2 &= \frac{1}{(\alpha-\beta)^2}[\beta d_1 + \alpha d_2], \\ \sigma_{22}^2 &= \frac{1}{(\alpha-\beta)^2}[d_1 + d_2]. \end{aligned}$$

Figure 1: Covariance restrictions and structural parameter space



Notes: In Panel a), the shaded area indicates the subset of the parameter space ruled out through the sign restriction imposed on α. In the remaining panels, the shaded area indicates the subset of the parameter space that cannot be reached by any structural representation of the model, given the value of ρ implicit in the covariance structure of the data. The lower the value of ρ, the more the data rule out positive values of β, pushing the posterior in the subset of the parameter space that supports the candidate restriction on β.

representation in the top-right quadrant of panel a. This means that 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 ρ is small or big would not be used. On the contrary, under a Bayesian perspective, the information on the magnitude of ρ can be used, because, for any prior distribution on α and β, the higher is ρ, the more the parameter space associated with the modes of the likelihood function overlaps with a region that contradicts the candidate identifying restrictions. This can be seen by comparing the figures in panels c and d, confronting the dark-shaded area and the light-shaded area. A statistical assessment can hence be derived, as long as the researcher is willing to exclude extreme values of α and β using a prior distribution that imposes zero mass in values considered too far away from zero. When excluding extreme values of the parameters, the probability mass of any prior distribution would be updated towards the top-right quadrant of panel a the more, the higher the correlation in the VAR innovations, hence suggesting a lower support of the data for the proposition analysed.

2.2 Simulation exercises

The discussion above showed that, while leaving the structural model unidentified, the reduced form model delivers information of where the indeterminacy region is along the parameter space. To help the exposition, the rest of the discussion proceeds using two simulation exercises.

Consider two alternative parametrization of model (1). For both parametrizations, I set

$$\alpha_{\text{true}} = 0.5 \;\; ; \;\; C(L)_{true} = 0 \;\; ; \;\; c_{1,\text{true}} = 1 \;\; ; \;\; c_{2,\text{true}} = 3 \;\; ; \;\; d_{1,\text{true}} = 0.1 \;\; ; \;\; d_{2,\text{true}} = 0.1.$$

I label the first equation as a supply function and the second equation as a demand function. Given the positive sign of α_{true}, the identifying restriction $\alpha > 0$ for the supply function is consistent with the model generating the data. However, the demand function would not always be correctly identified using $\beta < 0$. Consider two cases,

$$\beta_{true} = \begin{cases} -0.25 & \textit{Case 1} \\ +0.25 & \textit{Case 2} \end{cases}$$

Under *Case 1*, the demand function is downward sloping and would be correctly identified by the sign restriction $\beta < 0$. Under *Case 2* the demand function is upward sloping and cannot be identified with the same sign restriction used to identify the supply function.

The structural model can be written as

$$A_{true}\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{s}_t \quad , \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, D_{true}), \tag{4}$$

with

$$D_{true} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix} \quad ; \quad A_{true} = \begin{pmatrix} 1 & -\alpha_{\text{true}} \\ 1 & -\beta_{\text{true}} \end{pmatrix} = \begin{cases} \begin{pmatrix} 1 & -0.5 \\ 1 & 0.25 \end{pmatrix} & \textit{Case 1} \\ \begin{pmatrix} 1 & -0.5 \\ 1 & -0.25 \end{pmatrix} & \textit{Case 2} \end{cases}$$

The corresponding reduced form VAR representation $\boldsymbol{y}_t = A_{true}^{-1}\boldsymbol{c} + \boldsymbol{u}_t$ delivers a true covariance matrix Σ_{true} of the reduced form errors equal to

$$\Sigma_{true} = A_{true}^{-1} D_{true} A_{true}^{\prime -1} = \begin{cases} \begin{pmatrix} 0.0556 & 0.0444 \\ 0.0444 & 0.3556 \end{pmatrix} & \textit{Case 1} \\ \begin{pmatrix} 0.5000 & 1.2000 \\ 1.2000 & 3.2000 \end{pmatrix} & \textit{Case 2} \end{cases} \tag{5}$$

For the moment, simplify the analysis by starting the discussion in population. The covariance restrictions then take the form

$$\Sigma_{true} = A_i^{-1} D_i A_i'^{-1}, \tag{6}$$

with Σ_{true} given by (5). As a matter of illustration, the following are the Cholesky and to alternative candidate decompositions of Σ_{true} under *Case 1* and *Case 2*:

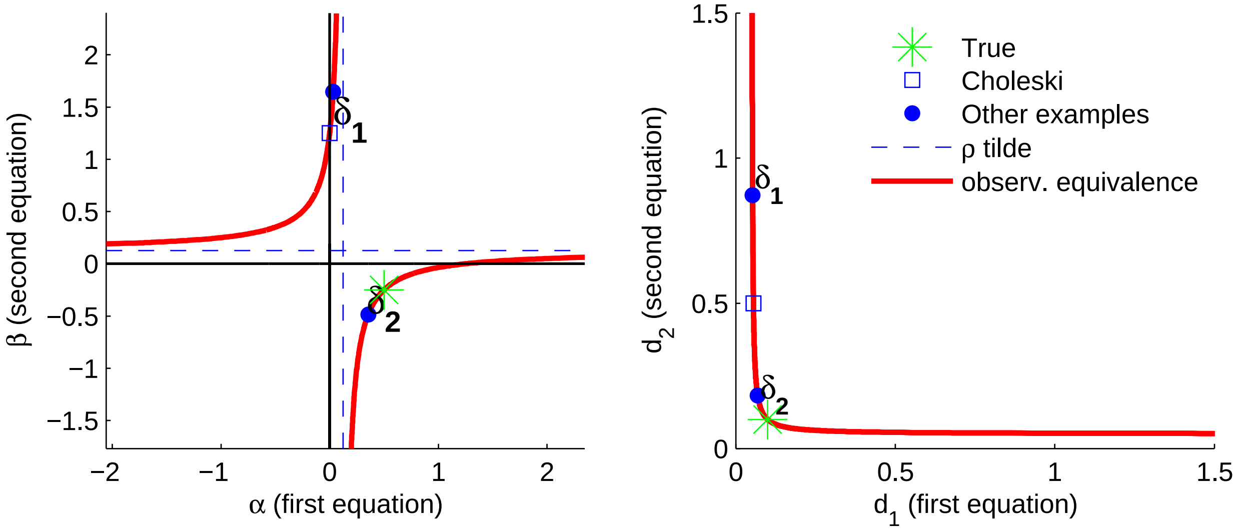
$$\textit{Case 1} \quad \boldsymbol{\delta}_{true} = \begin{pmatrix} \alpha \\ \beta \\ d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0.50 \\ -0.25 \\ 0.01 \\ 0.01 \end{pmatrix} \quad \boldsymbol{\delta}_{Chol} = \begin{pmatrix} 0 \\ 1.25 \\ 0.05 \\ 0.50 \end{pmatrix}, \ \boldsymbol{\delta}_1 = \begin{pmatrix} 0.03 \\ 1.64 \\ 0.05 \\ 0.87 \end{pmatrix}, \ \boldsymbol{\delta}_2 = \begin{pmatrix} 0.35 \\ -0.48 \\ 0.07 \\ 0.18 \end{pmatrix}$$

$$\textit{Case 2} \quad \boldsymbol{\delta}_{true} = \begin{pmatrix} \alpha \\ \beta \\ d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0.50 \\ 0.25 \\ 0.01 \\ 0.01 \end{pmatrix} \quad \boldsymbol{\delta}_{Chol} = \begin{pmatrix} 0 \\ 0.41 \\ 0.50 \\ 0.05 \end{pmatrix}, \ \boldsymbol{\delta}_1 = \begin{pmatrix} 0.32 \\ 0.66 \\ 0.06 \\ 0.31 \end{pmatrix}, \ \boldsymbol{\delta}_2 = \begin{pmatrix} 0.40 \\ -0.12 \\ 0.05 \\ 0.85 \end{pmatrix}$$

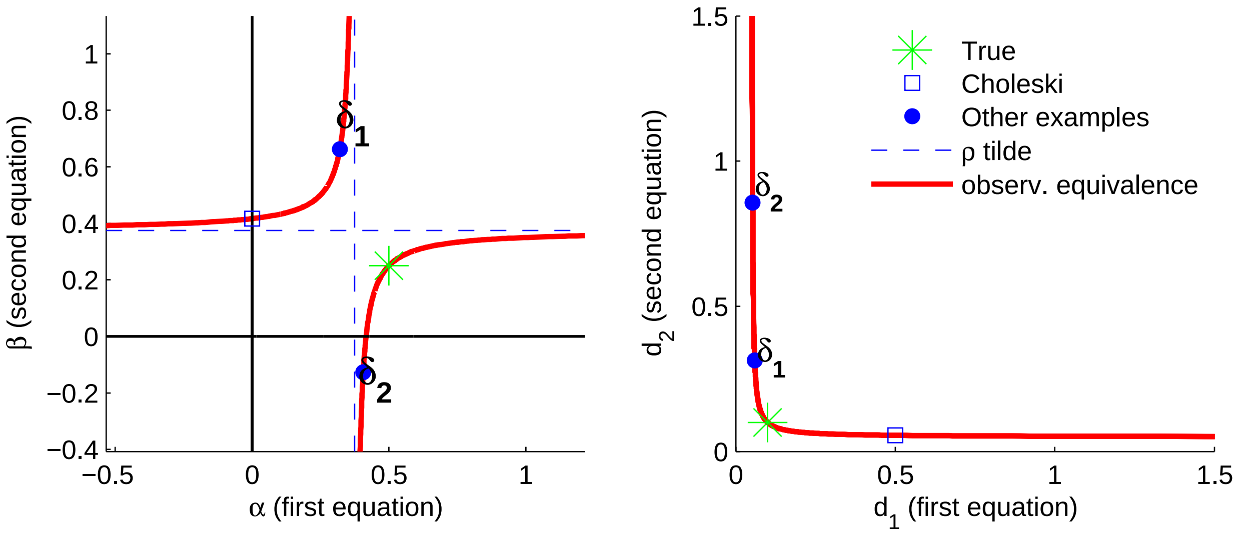
The covariance restrictions are analysed graphically in Figure 2. The line shows the set of observationally equivalent representations, i.e. combinations that are consistent with Σ_{true} (see the appendix for the details on the computation). The left graphs show combinations of α and β that satisfy the covariance restrictions, given an appropriate choice of d_1 and d_2. The right graphs show combinations of d_1 and d_2 that satisfy the covariance restrictions, given an appropriate choice of α and β. The green star reports the true combination of parameters. This lays along the set of observational equivalence because we have not introduced estimation uncertainty yet.

Figure 2: Set of observational equivalence

Case 1) candidate restrictions are *valid* in the model generating the data



Case 2) candidate restrictions are *invalid* in the model generating the data



Notes: The line indicates the set of observational equivalence of the model, consisting of all the structural representations that achieve the maximum value of the likelihood. The candidate representations $\boldsymbol{\delta}_{Chol}$ *and* $\{\boldsymbol{\delta}_i\}_{i=1}^2$ *reported numerically in the text are shown in the figures.*

Figure 2 clearly displays the nature of the identification problem of structural VARs. While not being completely flat, the likelihood function displays infinite modes along the line shown. 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. The modes of the likelihood are not distributed along the parameter space randomly, but they reflect the specific pattern of the covariance structure hidden in Σ_{true}.

As should be expected from the discussion in the previous section, the indeterminacy region for α and β is constrained to the region suggested in Figure 1, panels b, c, d. This admissible region is pinned down by the parameter ρ, which equals 0.1250 under *Case 1* and 0.3750 under *Case 1*. These values are reported in Figure 1 with the dashed lines. Under *Case 1* most of the mass of the indeterminacy region lays in the quadrants consistent with the candidate identifying restrictions. In contrast, under *Case 2* the indeterminacy region shifts towards the top right part of the parameter space, hence leading to a stronger mass of the posterior probability of any prior distribution towards the subset of the parameter space that contradicts the candidate identifying restrictions. Note that in both cases one can compute a structural representation consistent with the candidate identifying restrictions.

I now discuss the specification of the prior distribution used. Ideally, for the application in this paper, the prior distribution would be constructed in such a way as not to drive the Bayesian update, but to leave the update and the features of the posterior distribution to the likelihood function. In practice, a diffuse prior cannot be used because it would leave the marginal likelihood undefined, due to the indeterminacy of the model. Similarly, standard distributions like t-Student distributions for α and β and inverse Gamma distributions for d_1 and d_2 used for example in Baumeister & Hamilton (2015) would also be unsuitable for the purpose of this paper, because they would drive the update towards the parameter space mostly favoured by the prior. In this section I use independent uniform distributions for α, β, d_1, d_2. While being relatively uninformative, such prior implicitly imposes potentially unintended information on other features of the parameter (see for example the discussion in Giacomini & Kitagawa 2015). I will address this issue in Section 3.

It remains to specify the support of the prior distribution. In a real world application, a sufficiently wide range for this support can be constructed using either economic theory (as in Baumeister & Hamilton 2015, who calibrate the priors on α and β to imply 95% probability mass is in the support $[0, 2.2]$ and $[-2.2, 0]$, respectively), or using simple summary statistics from the data (as discussed in Section 4). In these simple simulations, I use a uniform distribution for β in the support $[-5, 5]$, which largely includes the true value in both cases considered. For α, in accordance with the proposition to be assessed, the support of the uniform prior distribution is $[0, 5]$, hence restricting the analysis to identifying α with sign restrictions. In all cases considered, the uniform prior distributions on d_1 and d_2 have support $[0, 2]$. Note that this equals up to 20 times the true value in the model generating the data.

The algorithm used to draw from the posterior distribution is discussed in the appendix. In short, for the update I use an estimate for Σ from a random sample of $T = 60$ observations, equivalent to 15 years of data in an hypothetical quarterly dataset. Given the above prior distribution, the posterior distribution coincides with the likelihood function, except that it equals zero in the parameter space in which the prior has zero mass. The posterior distribution does not have a standard shape and requires posterior simulation. The posterior sampler used is a Metropolis-Hastings-within-Gibb. The behaviour of the posterior distribuion can be appreciated by inspecting Figure 3. The posterior equals zero outside of the support of the prior distribution, and it features infinite modes along the infinite combinations along the set of observational equivalence of the likelihood. The posterior then decreases in correspondence to the parameter space which is off the set of observational equivalence.

Table 1: Bayesian assessment for the simulation exercises

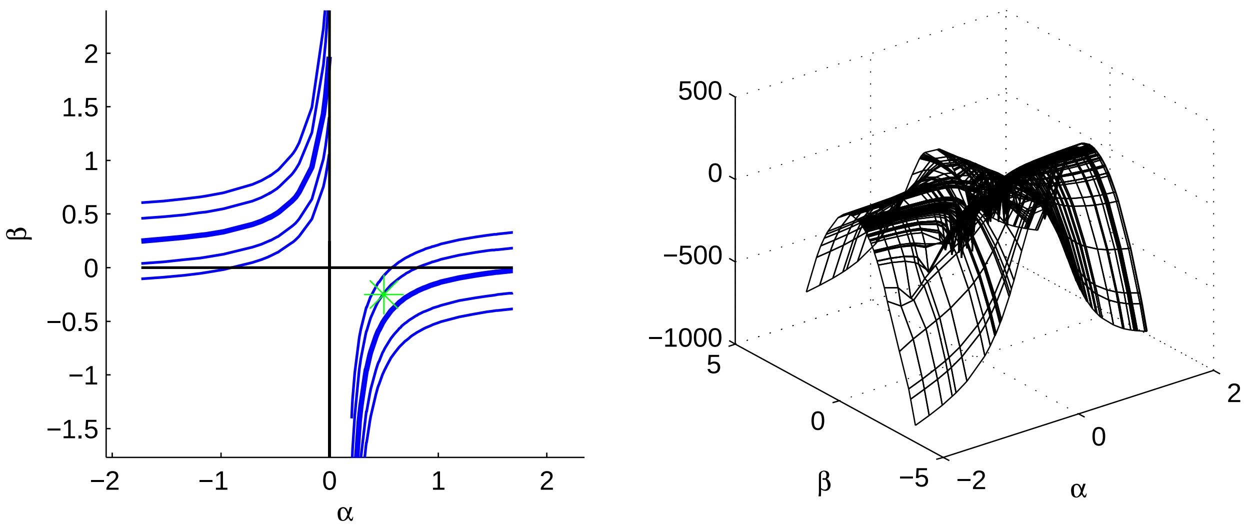
	Proposition assessed	Case studied	Satisfied in the DGP?	Prior probability in favour	Posterior probability in favour	Prior odds	Posterior odds
I	$\beta < 0 \mid \alpha > 0$	*Case 1*	yes	0.5	0.84	1	5.25
		Case 2	no	0.5	0.11	1	0.12
II	$\alpha > 0$ *and* $\beta < 0$	*Case 1*	yes	0.5	0.82	1	4.55
		Case 2	no	0.5	0.36	1	0.56
III	$\alpha < 0$ *and* $\beta < 0$	*Case 1*	no	0.25	0.11	0.12	0.89
		Case 2	no	0.25	0	0.33	0

The results of the analysis are shown in table 1. Consider first the proposition discussed so far. The researcher aims to assess if, conditioning on a positive sign of α, there is empirical support for β being negative. We know that this holds true for the model in *Case 1* and not true for the model in *Case 2*. The prior distribution discussed above yields 50% probability in favour and the remaining 50% probability against this proposition, and is the same for *Case 1* and *Case 2*. The corresponding prior odds equal 1. The Bayesian update delivers a posterior probability mass in favour of the proposition equal to 0.84 for *Case 1* and equal to 0.11 for *Case 2*, indeed correctly favouring the candidate identifying restrictions when such restrictions are consistent with the underlying model generating the data. The same can be seen from the corresponding posterior odds, which equal 5.25 for *Case 1* and 0.12 for *Case 2*.

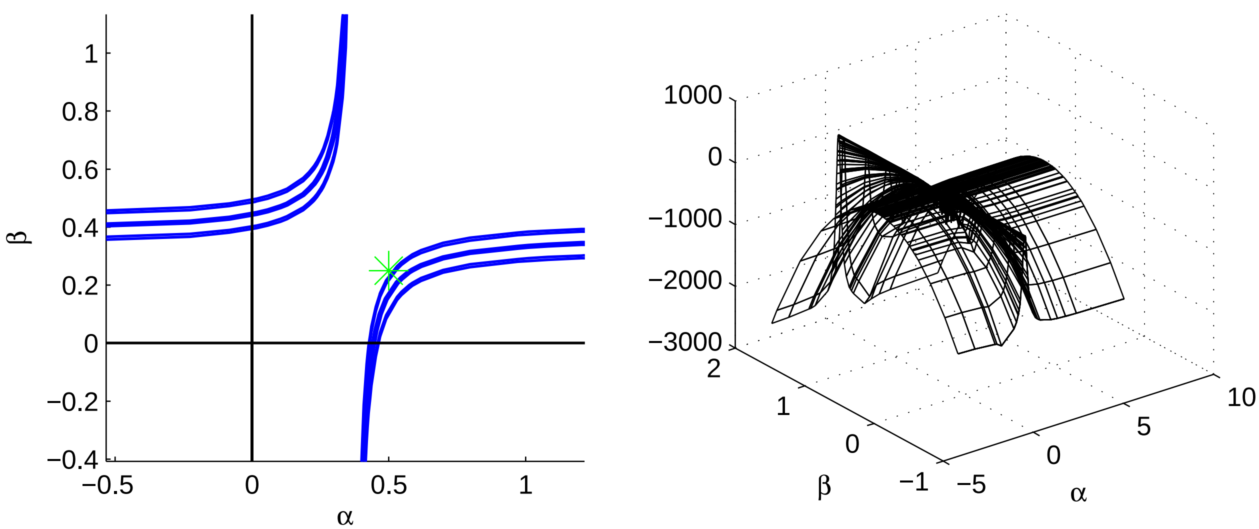
Consider now a different proposition, labelled in the table as II. The researcher aims to assess if the two equations of the model have slope of opposite sign. This differs from Proposition I for it does not condition on a restriction on α. We know

Figure 3: Likelihood function

Case 1) candidate restrictions are *valid* in the model generating the data



Case 2) candidate restrictions are *invalid* in the model generating the data



Notes: Compared to Figure 2, the true parameter values, indicated with the green star, do not lay on the maximum likelihood, due to estimation error.

that this holds true for the model in *Case 1* and not true for the model in *Case 2*. The prior distribution used is a uniform prior on the support $[-5, 5]$ for both α and β, and yields 50% probability in favour and the remaining 50% probability against this proposition. The prior odds are again equal to 1. The Bayesian update delivers a posterior probability mass in favour of the proposition equal to 0.82 for *Case 1* and equal to 0.36 for *Case 2*, indeed correctly favouring the candidate identifying restrictions when such restrictions are consistent with the model generating the data. Note that the inference becomes less precise, given that we are not conditioning on a correct restriction, namely $\alpha > 0$. The posterior odds equal 4.55 under *Case 1* and 0.56 under *Case 2*.

Last, consider Proposition III. The researcher aims to assess if both the equations of the model have negative slope. We know that this does not hold true in *Case 1* nor in *Case 2*. Given a positive correlation in both datasets, the frequentist approach would dismiss this possibility, as discussed in Figure 1. The Bayesian update delivers a posterior probability mass in favour of the proposition equal to 0.11 in the first case and to 0 in the second case, indeed correctly dismissing the proposition in both cases.

2.3 An application to Baumeister & Hamilton (2015)

I conclude this section by applying the algorithm to the analysis in Baumeister & Hamilton (2015). Baumeister & Hamilton (2015) estimate a structural VAR model on the labour market. As discussed above, they use the identifying restrictions that the wage elasticity of labour demand is negative and that the wage elasticity of labour supply is positive. They apply these identifying restrictions to the model using truncated t-Student prior distributions to the parameters indicating the elasticities of labour supply and labour demand. The Bayesian update developed in this paper allows to subjectively assess the validity of the identifying restrictions imposed.

As priors, I use independent uniform prior distributions in the support $[-5, 5]$ for the elasticities, and independent uniform prior distribution in the support $[0, 2]$ for variances of structural shocks. I apply the algorithm to the OLS-estimated variance covariance in their application (equation (38) in the original paper), using the length of their sample for T. When applying the algorithm, the probability mass favouring the proposition that the two equations of the structural model have slopes of opposite sign increases from 0.50 to 0.94 (i.e. prior odds of 1, posterior odds of 15.55), lending support to their identifying restrictions. When applying the algorithm to the propositions that the equations have slope of same sign the probability mass favouring such proposition decreases from 0.50 to 0.06 (i.e. prior odds of 1, posterior odds of 0.06), hence favouring again the identifying restrictions used. That this is the case can also be guess by comparing Figure 1 in this paper with Figure 4 in their paper.

3 Discussion and relation to the literature

Having outlined an illustrative example on the functioning of the Bayesian assessment proposed, I now discuss its main features and relate it to the literature. The next section discusses an application to the New Keynesian Model.

The fact that the data contribute to the Bayesian update despite the indeterminacy of the model can be appreciated as follows. Call Σ the covariance matrix of the VAR innovations, A_{Chol} the Cholesky decomposition of Σ such that $\Sigma = A^{-1}_{\text{Chol}} A'^{-1}_{\text{Chol}}$ and Q an orthogonal matrix. The parameter space spanned by the candidate structural representation corresponding to Σ can be inspected by drawing matrices Q and by generating matrices $A_i = Q \cdot A_{\text{Chol}}$. The likelihood function $L(\Sigma; data)$ of the model does not feature Q, implying that Q is unidentified.

Conditioning on Σ, no prior distribution $p(Q)$ would be revised upon observing the data, while any marginal revision would only be driven by a relationship between Σ and Q imposed with the prior, since Σ and Q are variation free, i.e. not constrained by one another (Poirier 1998, Dreze & Mouchart 1990). In this sense, the researcher does not learn about Q, unless external restrictions are introduced. However, A_{Chol} *is* identified, and the fact that A_{Chol} cannot then be associated with a unique matrix Q does not mean that A_i can be anywhere in the parameter space. As made clear from the analysis in Figure 1 and from the derivations in the appendix, the covariance structure of the model reveals whether certain parts of the parameter space can be reached or not, *for any matrix* Q, independently on its distribution, and independently on whether $P(Q)$ is revised or not. Hence, while not solving for the indeterminacy of the model, the researcher can form a posterior belief as to whether the structural representations lay in one part or another of the structural parameter space.

An equivalent way of expressing the above proposition is to partition the parametrization of the model into Σ and A_i. Given the likelihood function $L(\Sigma; data)$, A_i is unidentified. However, the parameter space for A_i is not variation free, since candidate A_i matrices are constrained by Σ through the covariance restrictions imposed by the model. Hence, upon learning about Σ, the researcher can use the constrains implicit in the covariance restrictions to learn where the candidate matrices A_i are in the parameter space. Canova & Sala (2009) and Koop et al. (2013) address this issue in the analysis of the identification of DSGE models.

Underidentification remains part of the model, as the Bayesian analysis developed in the paper does not aim to eliminate such feature of the model (see on this the discussion in Kociecki 2013). Such indeterminacy can be appreciated by the fact that, under the uniform prior used, the posterior distribution is equally high, say, in A_{Chol} and in QA_{Chol}, provided that they lay within the support of A. Graphically, this implies that the posterior distribution, say, in $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ in Figure 2, *Case 1*, is equally high. The Bayes factor would equal unity when comparing any two candidate representation that are equally consistent with Σ, as long as they belong to the support of the uniform prior of A. For this reason, the assessment developed is suitable only for set-identifying restrictions, not for point-identifying restrictions. It cannot, for instance, be used to assess if the data satisfy a recursive ordering of the structural shocks.

The need to preserve the non-identification of the model suggested the use of a uniform distribution as a prior distribution for A and D, since any other distribution could dominate the likelihood and deliver an unintended well-behave posterior. However, the uniform distribution is not neutral but it does impose restrictions in the update. The reasons, discussed for instance in Giacomini & Kitagawa (2015), relate to the fact that any uniform distribution of a parameter τ_1 implies a non-uniform distribution to parameter $\tau_2 = f(\tau_1)$ whenever $f(.)$ is a non linear function. This is indeed the case in VAR models, where impulse responses and other statistics of interest are complicated non linear functions of the parameters of the model. The paper does not aim at deriving completely uninformative priors, but to discuss the ability of the data to still identify something about the unidentified parameters of a model. See Giacomini & Kitagawa (2015) for a numerical algorithm that approximates extractions from posterior distributions independent from a specific prior distribution used to impose identifying zero and sign restrictions.

The proposed assessment of candidate set-identifying restrictions only provides the researcher with a probabilistic statement supporting the candidate identifying restrictions. Except for special cases, it will not rule out her uncertainty on whether the data generating process satisfies the restrictions or not. This can be appreciated from Table 1, where the posterior mass favouring Proposition I equalled 0.84, not 1. While reflecting the indeterminacy of the model (for example because the data under *Case 1* in Section 2.2 could have well be generated by the structural representation indicated in Figure 2 by $\boldsymbol{\delta}_1$), this feature is common in the use of posterior odds and Bayes factors in Bayesian model comparison. The update of subjective beliefs regarding the validity of the candidate restrictions then leaves the researcher with the choice of whether to rethink the identification strategy, or to trust the identifying restrictions and proceed with structural analysis, following for example Baumeister & Hamilton (2015).

Having discussed the features of the paper from a Bayesian perspective, I now relate the paper to frequentist uses of VAR models. In some applications of sign restrictions to VAR models, researchers report the acceptance ratio, defined as the ratio of structural candidate extractions that satisfy the sign restrictions relative to the total number of extractions (see for example Kilian & Murphy 2012 and Straub & Peersman 2006). The ultimate driver of the statistical assessment developed in the paper is the same feature that determines whether the researcher will find a high or a low acceptance ratio, although with two important differences.[2] In a frequentist setting, the only way of excluding a candidate set of sign restrictions is that the acceptance ratio equals exactly zero. For any value of the acceptance ratio strictly between zero and 1 the frequentist approach cannot reach a conclusion regarding the validity of the restrictions in the model generating the data. In the Bayesian approach developed in the paper, this acceptance ratio is mapped into a probability through a prior distribution, implying that a (subjective) probability interpretation can be

[2]Proposition II in Table 1 delivered a posterior in favour of the candidate restrictions equal to 0.84 under *Case 1* and 0.11 under *Case 2*. The acceptance ratio from 3000 extractions equals 0.78 and 0.19, respectively.

established. In addition, the acceptance ratio would fail to account for estimation uncertainty of the matrix Σ in the VAR, with the Bayesian approach accounts for the fact that Σ is not known in population. The latter point is also emphasized in Baumeister & Hamilton (2015).

Last, the analysis developed in the paper builds on the frequentist analysis by Paustian (2007). Paustian (2007) 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 that determines whether the candidate structural representations of the model will be in one part or another of the parameter space. I extent his analysis to a probabilistic setting, which allows to draw a conclusion 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 An application to the New Keynesian model

I conclude the analysis of the paper by showing an application of the Bayesian assessment procedure discussed so far to a simplified New Keynesian model. The model, which is used also in Koop et al. (2013), consists of three equations:

$$R_t = \psi \pi_t + \epsilon_t^{mp}, \tag{7}$$

$$x_t = E_t(x_{t+1}) - \sigma(R_t - E_t(\pi_{t+1})) + \epsilon_t^d, \tag{8}$$

$$\pi_t = \beta E_t(\pi_{t+1}) + \gamma x_t + \epsilon_t^s. \tag{9}$$

Equation (7) is a monetary policy rule determining the interest rate R_t. Equation (8) is an IS curve determining the output gap x_t. Equation (9) is a Phillips curve determining inflation π_t. The structural shocks in the model are a monetary shock ϵ^{mp}, a demand shock ϵ^d and a cost-push shock ϵ^s, or supply shock.

Rewriting the above model in structural VAR form gives

$$R_t = \psi\pi_t + \epsilon_t^{mp}, \tag{10}$$

$$x_t = -\sigma R_t + \epsilon_t^d, \tag{11}$$

$$\pi_t = \gamma x_t + \epsilon_t^s, \tag{12}$$

This, in turn, can be written as $A\boldsymbol{y}_t = \boldsymbol{\epsilon}_t$:

$$\underbrace{\begin{pmatrix} 1 & 0 & -\gamma \\ \sigma & 1 & 0 \\ 0 & -\psi & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} R_t \\ x_t \\ \pi_t \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} \epsilon_t^{mp} \\ \epsilon_t^d \\ \epsilon_t^s \end{pmatrix}}_{\boldsymbol{\epsilon}_t}. \tag{13}$$

I calibrate the parameter values of the data generating process as in Koop et al. (2013), setting $\sigma = 0.4$, $\gamma = 0.75$, $\psi = 2$. 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 sign identification of this type of models. Under the calibration used, the true reduced form model generating the data is

$$\underbrace{\begin{pmatrix} R_t \\ x_t \\ \pi_t \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} 0.62 & 0.93 & 0.47 \\ -0.25 & 0.62 & -0.18 \\ -0.50 & 1.25 & 0.62 \end{pmatrix}}_{B} \underbrace{\begin{pmatrix} \epsilon_t^{mp} \\ \epsilon_t^d \\ \epsilon_t^s \end{pmatrix}}_{\boldsymbol{\epsilon}_t}. \tag{14}$$

The variance of each structural shock is set equal to 1.5. $T = 60$ observations are generate from this model. The estimated variance-covariance matrix of the VAR innovations, which in this static model coincides with the estimated variance-covariance matrix of the data, is

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11}^2 & \hat{\sigma}_{21}^2 & \hat{\sigma}_{31}^2 \\ \hat{\sigma}_{21}^2 & \hat{\sigma}_{22}^2 & \hat{\sigma}_{23}^2 \\ \hat{\sigma}_{31}^2 & \hat{\sigma}_{32}^2 & \hat{\sigma}_{33}^2 \end{pmatrix} = \begin{pmatrix} 2.32 & 0.47 & 1.68 \\ 0.47 & 0.56 & 0.75 \\ 1.68 & 0.75 & 2.70 \end{pmatrix}. \tag{15}$$

Before developing the analysis, it is convenient to impose a normalization to the model in order to reduce the dimensionality of the parameter space considered. For simplicity, I normalize the first row of matrix B to $\bar{\beta}$, so that the remaining parameters in matrix B can be interpreted as the impact effect associated with a structural shock that increases the interest rate R_t by $\bar{\beta}$. The normalized matrix B is then

$$\tilde{B} = \begin{pmatrix} \bar{\beta} & \bar{\beta} & \bar{\beta} \\ \tilde{\beta}_{21} & \tilde{\beta}_{22} & \tilde{\beta}_{23} \\ \tilde{\beta}_{31} & \tilde{\beta}_{32} & \tilde{\beta}_{33} \end{pmatrix} \tag{16}$$

As in Section 2, this normalization allows to reduce the number of structural parameters that determine the validity of the restrictons assessed.

From system (14), the standard sign restrictions to identify a monetary shock, a demand shock and a supply shock are consistent with the data generating process, and hence they would be correctly imposed upon the data by a researcher. These restrictions are summarized as follows:

$$\text{Candidate sign restrictions} = \begin{bmatrix} \overbrace{+}^{\epsilon^{mp}} & \overbrace{+}^{\epsilon^{d}} & \overbrace{+}^{\epsilon^{s}} \\ - & + & - \\ \underbrace{-}_{\epsilon^{mp}} & \underbrace{+}_{\epsilon^{d}} & \underbrace{+}_{\epsilon^{s}} \end{bmatrix}.$$

The exercise developed in the rest of this section is the following. Suppose the researcher trusts the sign restrictions on a subset of shocks, but wants to assess whether she should impose the restrictions on the remaining shock(s). The Bayesian assessment developed in Section 2 can guide the researcher's subjective uncertainty regarding the validity of the remaining candidate identifying restrictions.

The researcher can set $\bar{\beta} = \hat{\sigma}_{11}$, which is the standard deviation of the interest rate R_t. The remaining parameters in (16) are hence interpreted as the effect of a structural shock of size such that the interest rate increases on impact by its standard deviation. The researcher can then set the prior distribution for $\tilde{b}_{ij}$ independently uniformly distributed in the range $[-5\hat{\sigma}_j^2, 5\hat{\sigma}_j^2]$. The intuition is that the researcher might want to constrain her subjective uncertainty regarding the structural parameters using available statistics from the reduced form model. In the case used, she imposes that any structural shock that moves the interest rate by its standard deviation cannot vary in the remaining variables by more than five times as much as their variances. This is intended to be a relatively wide support. When sign restrictions are imposed on some columns of $\tilde{B}$, the uniform distribution of the corresponding parameters is truncated to allow only for the desired signs.

The results of the exercise are shown in Table 2. Consider what is referred to in the table as *Scenario 1.* The researcher trusts the sign restrictions on the monetary shock and on the demand shock and wants to assess if the data support the candidate sign restrictions for a supply shock. She hence imposes the appropriate

Table 2: Bayesian assessment for the New Keynesian model

proposition assessed	Satisfied in the DGP?	Prior probability in favour	Posterior probability in favour	Prior odds	Posterior odds
Scenario 1					
Conditioning on ϵ^{mp} and ϵ^{d}, there is a ϵ^{s}	yes	0.25	0.97	0.33	40.07
Conditioning on ϵ^{mp} and ϵ^{d}, there is another ϵ^{mp}	no	0.25	0.00	0.33	0.00
Conditioning on ϵ^{mp} and ϵ^{d}, there is a ϵ^{d}	no	0.25	0.00	0.33	0.00
Conditioning on ϵ^{mp} and ϵ^{s}, "remaining option(s)"	no	0.25	0.02	0.33	0.02
Scenario 2					
Conditioning on ϵ^{mp} and ϵ^{s}, there is a ϵ^{d}	yes	0.25	0.37	0.33	0.60
Conditioning on ϵ^{mp} and ϵ^{s}, there is another ϵ^{mp}	no	0.25	0.00	0.33	0.00
Conditioning on ϵ^{mp} and ϵ^{s}, there is a ϵ^{s}	no	0.25	0.51	0.33	1.07
Conditioning on ϵ^{mp} and ϵ^{s}, "remaining option(s)"	no	0.25	0.09	0.33	0.10
Scenario 3					
Conditioning on ϵ^{d} and ϵ^{s}, there is a ϵ^{mp}	yes	0.25	0.38	0.33	0.61
Conditioning on ϵ^{d} and ϵ^{s}, there is another ϵ^{d}	no	0.25	0.36	0.33	0.57
Conditioning on ϵ^{d} and ϵ^{s}, there is another ϵ^{s}	no	0.25	0.04	0.33	0.05
Conditioning on ϵ^{d} and ϵ^{s}, "remaining option(s)"	no	0.25	0.20	0.33	0.26

truncated prior distributions for two of the columns of the $\tilde{B}$ matrix and uses a symmetric distribution for the remaining column. The truncated prior distribution used attaches probability in favour of such restriction equal to 0.25. We know that the assessed candidate restrictions are consistent with the data generating process and aim to find support for this proposition. Indeed, the posterior probability favouring the candidate identifying restrictions equals 0.97, corresponding to an increase from prior odds of 0.33 to posterior odds to 40.07. This shows correct support for the candidate identifying restrictions, reducing (but not eliminating) the researcher's subjective uncertainty on the validity of the additional restrictions.

Consider now the other cases shown in Table 2 for *Scenario 1*. These cases condition on the same sign restrictions for the monetary shock and the demand shock, and assess whether the data support the presence of another monetary shock, another supply shock and a shock featuring the residual possible signs (in this case, a vector featuring, in this order, a positive value and a negative value, hence not a monetary, nor a supply, nor a demand shock). We know that none of these cases hold true in the data generating process and we aim to find limited support for them.

Indeed, the update reduces the subjective probability attached to these options. The posterior odds equal essentially zero for all the three cases considered.

The remaining two Scenarios shown in Table 2 refer to the case in which the researcher conditions on the sign restrictions of a monetary shock and a supply shock (*Scenario 2*) or a demand shock and a supply shock (*Scenario 3*). The results show that the researcher successfully learns, for example, that there are two structural shocks with the same sign pattern associated with a monetary shock. In fact, conditioning on a monetary shock, the posterior odds favouring a second monetary shock equal 0. However, the algorithm struggles to distinguish between demand and supply shock, and attaches a high probability that there are either two supply shocks or two demand shocks. Overall, the researcher is confirmed in her identification of the monetary shock, but confirmed on the identification of a supply or of a demand shock only upon conditioning on a demand shock or a supply shock, respectively

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 assessing statistically the plausibility of identifying restrictions using the covariance restrictions. The analysis develops an approach which allows the data, together with a limited set of information, to speak in favour of or against candidate set-identifying restrictions, before such restrictions are imposed upon the data when actually identifying the model. What mainly drives the update is the correlation structure of the data. It is the correlation in the data that moves the mass of the prior distribution towards the part of the parameter space that is more consistent with the data generating process, given the normality assumption of the shocks and the constraints on extreme values imposed by the prior.

I show the functioning of the Bayesian update of set-identifying restrictions with two applications. The first one on a model of labour demand and supply, the second one in a New Keynesian Model featuring a monetary shock a supply shock and a demand shock. In both cases, the update provides useful information to the researcher, which is guided correctly in the update of her uncertainty regarding the restrictions that she is contemplating to use. At the same time, posterior uncertainty on the validity of the restrictions remains, and in some cases, the update fails to detect the true pattern of the data generating process.

References

Baumeister, C. & Hamilton, J. D. (2015), 'Sign restrictions, structural vector autoregressions, and useful prior information', *Econometrica* **83**(5), 1963–1999.

Canova, F. & Sala, L. (2009), 'Back to square one: identification issues in DSGE models', *Journal of Monetary Economics* **56**(4), 431–449.

Dreze, 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.

Dreze, J. H. & Mouchart, M. (1990), Tales of testing Bayesians, *in* R. A. Carter, J. Dutta & A. Ullah, eds, 'Contributions to Econometric Theory and Application', Springer.

Giacomini, R. & Kitagawa, T. (2015), 'Robust inference about partially identified SVARs'.

Hamilton, J. D. (1994), *Time series analysis*, Vol. 2, Princeton University Press.

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.

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.

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

The covariance restrictions of the model in Section 2 take the form

$$\Sigma = A^{-1} D A'^{-1}.$$

Calling $\sigma^2_{i,j}$ the i, j element of Σ, d_i the i, i element of D and under the following parametrization and normalization

$$A = \begin{pmatrix} 1 & -\alpha \\ 1 & -\beta \end{pmatrix},$$

the covariance restrictions can be written as

$$\begin{aligned} \sigma^2_{11} &= \frac{1}{(\alpha - \beta)^2}[\beta^2 d_1 + \alpha^2 d_2], \\ \sigma^2_{12} &= \frac{1}{(\alpha - \beta)^2}[\beta d_1 + \alpha d_2], \\ \sigma^2_{22} &= \frac{1}{(\alpha - \beta)^2}[d_1 + d_2]. \end{aligned}$$

This is a system of four unknowns $(\alpha, \beta, d_1, d_2)$ in three elements $(\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{22})$. Data (if working in sample) or population moments (if working in population) enter the restrictions through $(\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{22})'$, with $\sigma^2_{i,j}$ the i, j entry of Σ. Since these restrictions are non-linear in α and β, it is not immediately obvious what kind of information the covariance restrictions hide regarding the candidate solutions for α and β. Nevertheless, things become more clear if we rewrite the system recursively.

For this purpose, define $d = d_1 + d_2$ and $\omega = d_1/d$. We can rewrite the above system as

$$\frac{\sigma_{11}^2}{\sigma_{22}^2} = \beta^2 \cdot \omega + \alpha^2 \cdot (1-\omega),$$
$$\frac{\sigma_{12}^2}{\sigma_{22}^2} = \beta \cdot \omega + \alpha \cdot (1-\omega).$$

These two equations can be rearranged to isolate ω, obtaining

$$\omega = \frac{1}{\beta - \alpha}\Big(\frac{\sigma_{12}^2}{\sigma_{22}^2} - \alpha\Big), \tag{17}$$
$$\omega = \frac{1}{\beta^2 - \alpha^2}\Big(\frac{\sigma_{11}^2}{\sigma_{22}^2} - \alpha^2\Big). \tag{18}$$

Since by construction ω can only take values between 0 and 1, the above equalities imply a set of constraints on the candidate values of α and β. The first of these two equalities bears an economic intuition, which is discussed below. Substituting out ω, equations (17) and (18) determine a condition pinning down the solution for β given α:

$$\frac{\sigma_{11}^2}{\sigma_{22}^2} = \alpha^2 - \frac{\alpha^2 - \beta^2}{\alpha - \beta}\Big(\alpha - \frac{\sigma_{12}^2}{\sigma_{22}^2}\Big).$$

This equation implies that the set of observational equivalence shown in Figure 2 intersects the Cartesian axes for $(\alpha, \beta) = (0, \sigma_{11}^2/\sigma_{12}^2) = (\sigma_{11}^2/\sigma_{12}^2, 0)$. Given α and β, d is easily obtained using the last equality of the original system. The parameter ω can then be solved for using either equation (17) or equation (18).

Consider now equation (17). A necessary condition for a solution in α and β imposed by this equation

$$\text{Given } \rho = \frac{\sigma_{12}^2}{\sigma_{22}^2} = corr(r_q, r_p)\frac{\sigma_{11}}{\sigma_{22}}, \quad \text{if } \alpha \geq \rho \text{ then } \beta < \rho; \quad \text{if } \alpha \leq \rho \text{ then } \beta > \rho. \tag{19}$$

Not every combination of α and β that satisfies such condition is a candidate solution for system (6), it will be so only if it satisfies the full system.[3]

Appendix B: Metropolis-Hastings-within-Gibbs

The $n \times 1$ vector of reduced form shocks at time t is

$$\boldsymbol{r}_t \sim N(0, \Sigma),$$

with pdf

$$p(\boldsymbol{r}_t|\Sigma) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}\boldsymbol{r}_t'\Sigma^{-1}\boldsymbol{r}_t}.$$

Call R the $n \times T$ matrix gathering the observed or estimated T vectors of reduced form shocks $\boldsymbol{r}_t$. The data enter the likelihood through R. The likelihood function of R given Σ is

$$\begin{aligned} p(R|\Sigma) &= (2\pi)^{-\frac{Tn}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{r}_t'\Sigma^{-1}\boldsymbol{r}_t}, \\ &= (2\pi)^{-\frac{Tn}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{1}{2}\sum_{t=1}^{T}tr[\boldsymbol{r}_t'\Sigma^{-1}\boldsymbol{r}_t]}, \\ &= (2\pi)^{-\frac{Tn}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{1}{2}\sum_{t=1}^{T}tr[\boldsymbol{r}_t\boldsymbol{r}_t'\cdot\Sigma^{-1}]}, \\ &= (2\pi)^{-\frac{Tn}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{T}{2}tr\left[\sum_{t=1}^{T}\frac{\boldsymbol{r}_t\boldsymbol{r}_t'}{T}\cdot\Sigma^{-1}\right]}, \end{aligned}$$

where $tr(.)$ is the trace operator. This expression simplifies to

$$p(\hat{\Sigma}|\Sigma) = (2\pi)^{-\frac{Tn}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{T}{2}tr[\hat{\Sigma}\cdot\Sigma^{-1}]},$$

[3] Paustian uses a similar decomposition of the covariance restrictions and notes that, if we correctly impose that α is positive, then we will always correctly pick the negative sign of β under the condition that σ_{12}^2 is negative, i.e. under the condition that $d_{2,true}$ is large enough. I build on his argument and consider the information on α and β even if σ_{12}^2 is not negative, i.e. if his condition on the relative variances of the structural shocks is potentially not satisfied.

where $\hat{\Sigma} = \sum_{t=1}^{T} \frac{\boldsymbol{r}_t \boldsymbol{r}_t'}{T}$ is the OLS/MLE estimator of Σ. The data enter the likelihood through $\hat{\Sigma}$.

If we were using maximum likelihood estimation we would maximize the above expression in Σ, obtaining $\hat{\Sigma}$ (see Hamilton 1994). Instead, we aim for a Bayesian update of the structural components that underlie Σ. For this reason, rewrite the likelihood function in structural form by substituting

$$\Sigma = A^{-1} D A'^{-1}.$$

Using $|\Sigma| = |A^{-1}|^2 \cdot |D| = |A|^{-2} \cdot |D|$, the likelihood function becomes

$$p(\hat{\Sigma}|A, D) = (2\pi)^{-\frac{Tn}{2}} \cdot |D|^{-\frac{T}{2}} \cdot |A|^T \cdot e^{-\frac{T}{2} tr[\hat{\Sigma} \cdot A' D^{-1} A]}. \tag{20}$$

Last, remember that

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

Call $\boldsymbol{\gamma} = (\alpha, \beta)'$ and $\boldsymbol{d} = (d_1, d_2)'$. Then the likelihood function simplifies to a mapping of $\boldsymbol{\gamma}, \boldsymbol{d}$ into a real number using

$$p(\hat{\Sigma}|A, D).$$

As explained in Section 3.3, a natural candidate prior distribution for the analysis of this paper is an independent uniform distribution for each of the 4 elements in $\boldsymbol{\gamma}, \boldsymbol{d}$. Call S the support on which A, D have non-zero mass. The prior distribution is then

$$p(A, D) = \kappa \cdot I\{A, D \in S\},$$

where κ is a constant term that ensures that the mass under the prior distribution equals unity.

Having a prior distribution and a likelihood function, we can now derive the posterior distribution. Under a uniform prior, the posterior distribution is

$$p(A, D|\hat{\Sigma}) \propto I\{A, D \in S\} \cdot |D|^{-\frac{T}{2}} \cdot |A|^{T} \cdot e^{-\frac{T}{2}tr[\hat{\Sigma} \cdot A' D^{-1} A]}. \tag{21}$$

Put differently, the posterior distribution behaves in exactly the same way as the likelihood, except that it equals zero if any of the variables in $\boldsymbol{\gamma}, \boldsymbol{d}$ are outside the support S on which the prior has positive mass.

The posterior distribution is clearly not a standard distribution, and we need to numerical approach to draw from it. Nevertheless, before developing the algorithm, it is convenient to inspect how the posterior distribution behaves. The key feature of (21) is that it displays infinitely many modes, and this must be taken into account when developing the numerical algorithm. This is simply a consequence of the fact that there are infinite combinations of $\boldsymbol{\gamma}, \boldsymbol{d}$ which achieve the highest point of the likelihood, as outlined in Section 3.2.

The joint posterior distribution of equation (21) behaves in such a way as to suggest a Metropolis-Hastings-within-Gibbs algorithm as a natural algorithm. In fact, while the joint posterior distribution of $\boldsymbol{\gamma}, \boldsymbol{d}$ is not an immediate object to analyse, the conditional posterior distributions $p(\boldsymbol{\gamma}|\boldsymbol{d}, \hat{\Sigma})$ and $p(\boldsymbol{d}|\boldsymbol{\gamma}, \hat{\Sigma})$ are easier. Using the notation from Section 3.2, call $\boldsymbol{\delta} = (\boldsymbol{\gamma}', \boldsymbol{d}')'$. Take $\boldsymbol{\delta_i}$ as one element of the set of observational equivalence. Consider $p(\boldsymbol{\gamma}|\boldsymbol{d_{\delta_1}}, \hat{\Sigma})$ where $\boldsymbol{d_{\delta_1}}$ is the set of values of d_1 and d_2 corresponding to the combination $\boldsymbol{\delta_1}$. We can see from Figure 2 that, conditioning on d_1 and d_2 belonging to $\boldsymbol{\delta_1}$, the posterior distribution of $\boldsymbol{\gamma}$ should have a mode on the dot indicated in the top graph of the figure with $\boldsymbol{\delta_1}$. Similarly, consider $p(\boldsymbol{d}|\boldsymbol{\gamma_{\delta_2}}, \hat{\Sigma})$ where $\boldsymbol{\gamma_{\delta_2}}$ is the set of values of α and β corresponding to the combination $\boldsymbol{\delta_2}$ of the observational equivalence set. We can see from Figure 2, bottom graph, that, conditioning on α, β belonging to $\boldsymbol{\delta_2}$, the posterior distribution of $\boldsymbol{d}$ should have a mode on the dot indicated in the figure with δ_2.[4] Put it differently, while the joint posterior distribution features infinitely many modes, the conditional posterior distributions do not, and can be analysed naturally with a Metropolis-Hastings algorithm. The Metropolis-Hastings algorithms for the conditional distributions can then be organized into a Gibbs sampler in order to ensure that the full algorithm samples all parts of the indeterminacy region.

More precisely, I use the following algorithm to sample from the posterior distribution of equation (21):

1. generate $m_1 = 50$ random extractions of candidate decompositions $\boldsymbol{\delta_i}$ of $\hat{\Sigma}$. This set gives an approximate indication of where the indeterminacy region is across the parameter space. For each of these extractions run the following algorithm;

2. take the parameter values for $\boldsymbol{d}$ corresponding to $\boldsymbol{\delta_i}$. Conditioning on such values of $\boldsymbol{d}$, use numerical maximization to find a mode of $\boldsymbol{\gamma}|\boldsymbol{d}, \hat{\Sigma}$ using as an initial guess any random decomposition belonging to the set of observational equivalence. Use a Metropolis Hastings algorithm starting from the mode of $\boldsymbol{\gamma}$ to explore the distribution of $\boldsymbol{\gamma}|\boldsymbol{d}, \hat{\Sigma}$. The algorithm uses a random walk of the type

$$\boldsymbol{\gamma}_{new} = \boldsymbol{\gamma}_{old} + \tau \cdot P \cdot \boldsymbol{u},$$

[4]More precisely, conditioning on $\boldsymbol{\gamma}$ (or on $\boldsymbol{d}$), $\boldsymbol{d}$ (or $\boldsymbol{\gamma}$) has two modes. This is not an issue when using the algorithm outlined below.

with $\boldsymbol{u}$ the realization of a 2×1 $t-student$ with 4 degrees of freedom and P the Cholesky factorization of the Hessian matrix of the conditional pdf evaluated in the mode of $\boldsymbol{\gamma}$. The tuning parameter τ will be set to achieve an average acceptance probability of around 50%. Run the algorithm $m_2 = 100$ times. Keep and store only the last extraction of $\boldsymbol{\gamma}$ and call it $\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}$. This essentially burns-in the first $m_2 - 1$ observations and allows the Gibbs algorithm to proceed conditioning on a single vector of observations;

3. Conditioning on $\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}$, inspect the posterior distribution of $\boldsymbol{d}$ using the same procedure above, i.e. running the Metropolis-Hastings algorithm for $m_2 = 100$ iterations. Then keep only the last extraction, defined as $\bar{\boldsymbol{d}}_{\boldsymbol{i}}$;

4. Run the algorithm $m_3 = 100$ times and store the corresponding extractions for $\boldsymbol{a}, \boldsymbol{d}$. Combined with the loop of point 1, this gives $m_1 \times m_3 = 5000$ extractions.