Sign restrictions in high-dimensional vector autoregressions

Dimitris Korobilis
University of Glasgow

September 14, 2020

Abstract

This paper proposes a new Bayesian sampling scheme for inference in vector autoregressions (VARs) using sign restrictions. I build on a factor model decomposition of the reduced-form VAR disturbances, which are assumed to be driven by a few common factors/shocks. The outcome is a computationally efficient algorithm that allows to jointly sample VAR parameters as well as decompositions of the covariance matrix satisfying desired sign restrictions. Using artificial and real data I show that the new algorithm works well and is multiple times more efficient than existing accept/reject algorithms for sign restrictions.

Keywords: high-dimensional VAR; structural inference; factor model; sign restriction; Gibbs sampling

JEL Classification: C11, C13, C15, C22, C52, C53, C61

This paper is partial outcome of work I undertook as an external expert for the Monetary Analysis Division of the European Central Bank (ECB). The views expressed in the paper are solely mine (as are any remaining errors) and they do not necessarily reflect views of the ECB or the Eurosystem.

I would like to thank without implicating Christiane Baumeister, Martin Bruns, Fabio Canova, Filippo Ferroni, Luca Gambetti, Gary Koop, Michele Lenza, Alberto Musso, Serena Ng, Michele Piffer, Davide Pettenuzzo, John Tsoukalas and Harald Uhlig, as well as numerous seminar and conference participants, for useful discussions and comments.

Correspondence: Professor of Econometrics, Department of Economics, University of Glasgow, 40 University Avenue, Glasgow, G12 8QQ, UK; email:dikorobilis@googlemail.com

1 Introduction

This paper proposes a new Bayesian Markov chain Monte Carlo (MCMC) algorithm for joint estimation of parameters of reduced-form vector autoregressions (VARs) and associated sign restrictions for structural identification. The main idea is to allow the reduced-form VAR disturbances to have a static factor model structure. By doing so, sign and other restrictions can be incorporated via straightforward parametric prior distributions; the factors can be interpreted as structural VAR disturbances; and the VAR likelihood can be used to test the plausibility of identification restrictions as well as the general fit of the VAR. Most importantly, the specific factor formulation of the VAR disturbance terms allows for the derivation of a simple and highly efficient Gibbs sampling scheme that jointly samples VAR parameters and sign restrictions. Consequently, the main contribution of this paper is to establish a comprehensive methodology for estimation and identification in VARs that is computationally efficient, to the extent that it can be used with an arbitrarily large number of endogenous variables and/or shocks.

The sign restrictions approach to identification has become increasingly popular in applied work relative to traditional identification methods, such as exclusion restrictions; see Kilian and Lütkepohl (2017) for a detailed review of this literature. The main feature of existing Bayesian algorithms for inference in sign restrictions, such as Rubio-Ramírez et al. (2010) and Baumeister and Hamilton (2015), is that they rely on rejection sampling schemes (also known as *accept/reject algorithms*) in order to search for matrices that satisfy the desired restrictions. If restrictions are tight, as it would be the case in models with many variables and many shocks, rejection sampling results in constantly rejecting draws. In contrast, the Gibbs sampler proposed in this paper allows to sample sign-restricted matrices of contemporaneous structural relations from their conditional posterior and these samples are always accepted. The benefits of the new approach are demonstrated in a reasonably-large VAR using 15 variables for the US. Using synthetic data I show that one can push the VAR dimension to much larger values, as it takes only 11 minutes in a standard modern personal computer to estimate a 100-variable

VAR(1) with 20 shocks and 200 time series observations.[1] After establishing that the new algorithm is very fast, the main purpose of this paper is to establish that the algorithm is also numerically sensible. While the new algorithm is derived from a specific factor-VAR methodology, it is shown that its output is qualitatively quite similar to the output of the algorithms of Rubio-Ramírez et al. (2010) and Arias et al. (2018) that are based on traditional reduced-form VARs.

The proposed factor-VAR methodology is related to previous attempts and efforts for structural VAR identification.[2] Gorodnichenko (2005) specified an identical VAR model with reduced-rank decomposition of the disturbance terms. However, he used this specification in order to replace standard block diagonal restrictions in VARs (Bernanke and Blinder, 1992), with a more parsimonious identification scheme that imposes less (possibly unreasonable) zero restrictions. More recently, Matthes and Schwartzman (2019) specify a closely related VAR model in order to identify the structural impact of sectoral dynamics on GDP. Their identification is via a factor structure on the residuals that has the additional assumption of allowing for correlation within industries but no correlation across industries.

Similarly, Stock and Watson (2005a) specify a more general factor-augmented VAR (FAVAR) and discuss in detail how various identification schemes fit in this setting. They also note (Stock and Watson, 2005a, Section 3.5) that the sign restrictions identification scheme proposed by Uhlig (2005) also fits the FAVAR framework. An application of this idea can be found in Ahmadi and Uhlig (2015). From a modeling point of view, the factor model I propose can be viewed as a special case of the Ahmadi and Uhlig (2015) FAVAR. However, the specification proposed in this paper has completely different implications both

[1] Fair timing comparisons between the new algorithm and existing algorithms are hard to set up, as these will be affected by many factors (starting from modeling choices, such as the number of restrictions imposed, to other factors such as the programming language used). As a rough indication, I find in the empirical section that obtaining 5,000 draws from the benchmark six-variable VAR of Furlanetto et al. (forthcoming) using the new algorithm takes less than five minutes; using the original Rubio-Ramírez et al. (2010) algorithm that Furlanetto et al. (forthcoming) adopted in order to produce their results, it takes roughly four hours to sample 2,000 models that satisfy the same restrictions.

[2] To be exact, the VAR with factor structure in the disturbances is inspired by the panel data literature, and such structure has been used before in multi-country VARs; see Stock and Watson (2005b).

algorithmically and in terms of inference. Ahmadi and Uhlig (2015) project a large vector of observable macroeconomic variables into a smaller vector of factors and they model VAR dynamics only on these factors. This means that there is some loss of information (not all macro variables are explained well by the factors) and the statistical fit of the factors determines the contribution of each structural shock on each macroeconomic variable. Additionally, the autoregressive dynamics of the large macro dataset is represented only by the autoregressive dynamics of the smaller vector of factors. This modeling choice means that, inevitably, the FAVAR is unable to capture richer patterns of propagation of structural shocks to observed macroeconomic variables. In contrast, in this paper all observable macroeconomic variables are endogenous in the VAR and the sole role of the factors is to represent structural shocks. Additionally, the proposed algorithm is computationally simpler as it relies on posterior formulas for linear regression models, instead of building on more demanding simulation smoothing techniques, as is the case with the FAVAR (see Ahmadi and Uhlig (2015) and Bernanke et al. (2005)).

In the next Section I introduce the new methodology and associated Gibbs sampler algorithm for inference, and I outline the key components that help speed up and stabilize (numerically) posterior sampling in high dimensions. In Sections 3 and 4 I undertake several important exercises using artificial and real datasets, in order to illustrate the excellent numerical properties of the new algorithm. Section 5 concludes the paper.

2 VARs driven by a few, common shocks

The starting point is the following reduced-form vector autoregression

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\mathbf{y}_t$ is a $(n \times 1)$ vector of observed variables, $\mathbf{x}_t = \left(1, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-p}\right)'$ a $(k \times 1)$ vector (with $k = np+1$) containing a constant and p lags of $\mathbf{y}$, $\mathbf{\Phi}$ is an $(n \times k)$ matrix of coefficients,

and $\mathbf{u}_t$ a $(n \times 1)$ vector of structural disturbances assumed to be $N(\mathbf{0}_{n\times 1}, \mathbf{\Omega})$ with $\mathbf{\Omega}$ an $n \times n$ covariance matrix.

The most computationally efficient algorithm for sampling sign restrictions in medium-large Bayesian VARs is developed by Rubio-Ramírez et al. (2010) and extended by Arias et al. (2018).[3] This algorithm can be summarized using the following steps (Kilian and Lütkepohl, 2017):

1. Use posterior sampling to obtain R samples from the joint posterior of $(\mathbf{\Phi}, \mathbf{\Omega})$.
2. Take a single draw $(\mathbf{\Phi}^r, \mathbf{\Omega}^r)$, $r = 1, ..., R$, and compute the Cholesky factor $\mathbf{P}^r = chol(\mathbf{\Omega}^r)$.
3. Obtain S draws of an orthogonal rotation matrix $\mathbf{Q}^s$, $s = 1, ..., S$, and compute $\widetilde{\mathbf{A}} = \mathbf{P}^r\mathbf{Q}^s$, where $\widetilde{\mathbf{A}}$ satisfies $\widetilde{\mathbf{A}}\widetilde{\mathbf{A}}' = \mathbf{\Omega}^r$.
4. If $\widetilde{\mathbf{A}}$ satisfies the desired sign restrictions retain and store this draw, otherwise discard it.
5. Repeat steps 2,3,4 above R times.

Regarding step 1, there is a large recent literature on how to obtain samples from the posterior of high-dimensional Bayesian VARs. Step 2 relies on simply taking the Cholesky decomposition, while for step 3 Rubio-Ramírez et al. (2010) show that one can generate a matrix $\mathbf{W}$ from $N(0, I)$ distribution and then use its QR decomposition to obtain $\mathbf{Q}$. Both the Cholesky and QR factorizations have $\mathcal{O}(n^3)$ algorithmic complexity (number of floating points), such that these operations could only become computationally cumbersome in the improbable scenario that a researcher wants to estimate a VAR with thousands of endogenous variables. The basic algorithm above can be further improved by noting that it is trivial to parallelize (hence, use modern processing capabilities), and that the same S rotation matrices $\mathbf{Q}$ can be used for each of the R samples of $(\mathbf{\Phi}, \mathbf{\Omega})$.

[3] Other algorithms exist, such as the Metropolis-Hastings algorithm for SVARs of Baumeister and Hamilton (2015), the penalty-function approach (PFA) of Mountford and Uhlig (2009) and the accept-reject algorithm of Ouliaris and Pagan (2016). As Arias et al. (2018) note, the Metropolis-Hastings algorithm of Baumeister and Hamilton (2015) becomes inefficient even in VARs of medium size. The Ouliaris and Pagan (2016) algorithm was developed for inference partially identified VAR models estimated with least squares, and it is not clear how it would generalize to VARs with thousands of parameters. Finally, the PFA approach has been shown to imply additional unintended sign restrictions; see the discussion in Arias et al. (2018) and (Kilian and Lütkepohl, 2017, Section 13.6.4).

Nonetheless, the real culprit that leads this algorithm to fail in high dimensions is the accept/reject step 4. Rejection sampling has been very popular for sampling from distributions since the early work of von Neumann (1951), and it is a well-established that such sampling scheme will fail if the function that is sampled is highly concentrated in a certain region of its support. An intuitive example why this will be the case in large dimensions is the following: Assume that we are interested in a VAR with 50 variables and we want to identify using sign restrictions 20 shocks that we believe they describe our economy of interest. Without loss of generality, assume we impose sign restrictions in all 50 variables over these 20 shocks. What is the probability that we can obtain a sample from the posterior of $\mathbf{P}$, generate randomly a rotation matrix $\mathbf{Q}$, and find that all 1000 restrictions in $\widetilde{\mathbf{A}} = \mathbf{P}^r\mathbf{Q}^s$ are satisfied? The answer is that this probability becomes virtually zero. It becomes apparent that in high dimensions the desired restrictions will be so tight that no sample of $\widetilde{\mathbf{A}}$ will be accepted, see also the discussion in Section 13.6.4 of Kilian and Lütkepohl (2017).

The solution proposed here is to generate a matrix $\widetilde{\mathbf{A}}$ that satisfies the required sign restrictions by using a scheme where every proposed sample is accepted. In order to achieve this aim, the formulation of the VAR has to be modified in order to consider inference on $\left(\mathbf{\Phi}, \mathbf{\Omega}, \widetilde{\mathbf{A}}\right)$ as a joint estimation problem, rather than first estimate a VAR and then try to identify sign restrictions in two steps.

The proposed solution builds on fundamental ideas in the factor model literature, as applied to empirical problems in macroeconomics: a few common forces (which in a structural setting we desire to identify as "primitive shocks"; see Ramey, 2016) are driving the set of shocks to a system of n endogenous variables. In order to materialize this idea, the reduced-form VAR disturbances in equation (1) are decomposed using the following static factor form

$$\boldsymbol{\varepsilon}_t = \mathbf{\Lambda}\mathbf{f}_t + \mathbf{v}_t, \tag{2}$$

where $\mathbf{\Lambda}$ is an $n \times r$ matrix of factor loadings, $\mathbf{f}_t$ is an $r \times 1$ vector of factors, and $\mathbf{v}_t$ is an

$n \times 1$ vector of idiosyncratic shocks. The crucial assumption here is that n is large and that $r < n$ (and not necessarily $r \ll n$, as is typically assumed in the factor literature). In line with the "*exact factor model*" literature, let $\mathbf{v}_t \overset{\text{i.i.d}}{\sim} N(\mathbf{0}_{n\times 1}, \mathbf{\Sigma})$, with $\mathbf{\Sigma}$ a diagonal matrix. Additionally, let $\mathbf{f}_t \sim N(\mathbf{0}_{r\times 1}, \mathbf{I}_r)$, which means that the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ is now of the form

$$cov(\boldsymbol{\varepsilon}_t | \mathbf{\Lambda}, \mathbf{\Sigma}) = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Sigma}. \tag{3}$$

This factor model decomposition of $\mathbf{\Omega}$ shows that, as long as $\mathbf{\Sigma}$ is diagonal, identification via sign restrictions can be achieved by imposing the desired signs on $\mathbf{\Lambda}$. To see this consider again a reduced-rank structural VAR representation of this model, which can be obtained by multiplying the reduced-form VAR model implied by equations (1) - (2) with the generalized inverse of $\mathbf{\Lambda}$, as follows:

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{\Phi}\mathbf{x}_t + \mathbf{\Lambda}\mathbf{f}_t + \mathbf{v}_t & (4)\\
(\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{y}_t &= (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{\Phi}\mathbf{x}_t + \mathbf{f}_t + (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{v}_t & (5)\\
\mathbf{A}_1\mathbf{y}_t &= \mathbf{B}_1\mathbf{x}_t + \mathbf{f}_t + (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{v}_t. & (6)
\end{aligned}$$

In the equation above, the structural VAR matrix $\mathbf{A}_1$ is equivalent to the generalized inverse $(\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'$. While $\mathbf{\Lambda}$ is not observed, assume for a moment that a consistent estimator of this parameter exists. Given that in the exact factor model formulation the $\mathbf{v}_t$ are uncorrelated, the CLT in Bai (2003) suggests that for each t and for $n \to \infty$ we have that $(\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{v}_t \to 0$ making this term asymptotically negligible.[4] Therefore, $\mathbf{f}_t$ in the SVAR equation (6) can be seen as a projection of the structural disturbances $\mathbf{u}_t$ into a lower dimensional space $\mathbb{R}^r$.

Finally, it should be noted that the proposed specification has a drawback, as it only allows to impose sign restrictions upon impact of a shock, and not carry on the sign restrictions into subsequent periods. In practice, there is little consensus in economic theory about the signs of

[4]Note that in the proposed algorithm the impact of $\mathbf{v}_t$ is not neglected, since I work exclusively with the reduced VAR form. This result is here only to demonstrate that, as the dimension of the VAR increases, the factors can be viewed as being structural disturbances.

structural impulse responses at longer horizons (see for example Canova and Paustian, 2011), and for that reason the vast majority of empirical papers do impose restrictions on impact (Kilian and Lütkepohl, 2017). Nevertheless, as Uhlig (2017) notes, it can be quite useful to have the option to impose sign restrictions in the long-run response of macroeconomic variables to shocks. Within the context of the proposed methodology, this issue can be addressed if equation (2) is specified as a dynamic factor model instead of a static factor model. However, this more general assumption would require to rely on filtering and smoothing sampling steps that would lead to a completely different algorithm overall compared to the algorithm presented in this paper. As a result, I exclusively focus here on the problem of how to impose impact sign restrictions in high-dimensional VARs. Extending to the case of long-horizon sign restrictions is feasible, but it is left for future research.

2.1 A new Gibbs sampler for sign restrictions in reduced-form VARs

Before deriving a Gibbs sampler algorithm for the high-dimensional VAR with factor structure in the residuals, identification issues have to be clarified. Following Anderson and Rubin (1956), notice that $\mathbf{\Omega}$ has $n(n+1)/2$ free parameters while the factor decomposition in equation (3) has $nr+n$ free parameters. Therefore, the first condition for estimation is that $n(n+1)/2 \geq nr + n$ or that $r \leq (n-1)/2$. This condition implies that – not taking into account any zero or sign restrictions on $\mathbf{\Lambda}$ – in a 19-variable VAR a reasonable number of nine factors/shocks can be estimated. In practice, any zero or sign restrictions we impose on $\mathbf{\Lambda}$ (depending on the empirical application) will allow to lift, to a large extent, this upper bound on the number of factors. Another issue in the factor decomposition is how $\mathbf{\Lambda}$ and $\mathbf{f}_t$ can be identified jointly from the likelihood. As both these matrices are latent, an estimate $\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{f}}_t$ can be rotated to an observationally equivalent solution using any $r \times r$ orthogonal matrix. Lopes and West (2004) discuss this issue in detail and they follow the typical approach in identification of Bayesian factor models, which is to impose zero restrictions on the upper $r \times r$ block of $\mathbf{\Lambda}$. Such

restrictions are driven from the desire for statistical – not structural – identification, and are usually suboptimal since they hardly ever fit real data precisely.[5] Instead, in the factor-VAR model the factors $\mathbf{f}_t$ are equivalent to structural shocks that are uncorrelated. Therefore, I restrict their posterior to be $N(\mathbf{0}_r, \mathbf{I}_r)$. Under the assumption that $\mathbf{\Lambda}$ will also have additional sign or zero restrictions, the structural shocks (factors) will be identified. However, even in cases where we do not have enough sign restrictions for identification of the factors, we can still fully identify the common component $\mathbf{\Lambda f}_t$ as well as the decomposition of $\mathbf{\Omega}$ in equation (3). That is, we are always able to sample a decomposition of $\mathbf{\Omega}$ that embeds the desired sign restrictions, even when the structural shocks are not fully identified (as long as the structural shocks are uncorrelated and normalized to have variances equal to one).

Posterior sampling in the reduced-form VAR with factor structure in the residuals is straightforward using the Gibbs sampler. This is because posterior conditional distributions have very standard forms and are trivial to sample from. To see this, write the model using a single equation for convenience

$$\mathbf{y}_t = \mathbf{\Phi x}_t + \mathbf{\Lambda f}_t + \mathbf{v}_t. \tag{7}$$

Assume that all sign and zero restrictions on $\mathbf{\Lambda}$ are collected into a matrix $\mathbf{S}$, with entries $+1$ for positive signs, -1 for negative signs, 0 for zero restrictions, and a missing value for no restriction (we denote this case in this paper using the symbol NA, and in the code using the

[5] For example, one issue with restrictions used in Bayesian factor models, such as Lopes and West (2004) or Bernanke et al. (2005), is that the ordering of the variables in $\mathbf{y}$ plays a role in estimation of the factors.

MATLAB value NaN). The priors for the VAR parameters are of the form

$$\begin{aligned}
\boldsymbol{\phi}_i \equiv vec\left(\boldsymbol{\Phi}_i\right) &\sim N_k\left(\mathbf{0}, \underline{\mathbf{V}}_i\right), && (8)\\
\mathbf{f}_t &\sim N_r\left(\mathbf{0}, \mathbf{I}\right), && (9)\\
\boldsymbol{\Lambda}_{ij} &\sim \begin{cases} N\left(0, \underline{h}_{ij}\right) I(\Lambda_{ij} > 0), & \text{if } S_{ij} = 1, \\ N\left(0, \underline{h}_{ij}\right) I(\Lambda_{ij} < 0), & \text{if } S_{ij} = -1, \\ \delta_0\left(\boldsymbol{\Lambda}_{ij}\right), & \text{if } S_{ij} = 0, \\ N\left(0, \underline{h}_{ij}\right), & \text{otherwise,} \end{cases} && (10)\\
\sigma_i^2 &\sim inv - Gamma\left(\underline{\rho}_i, \underline{\kappa}_i\right), && (11)
\end{aligned}$$

for $i = 1, ..., n$, $j = 1, ..., r$, where $\boldsymbol{\Phi}_i$ is the i^{th} row of $\boldsymbol{\Phi}$, σ_i^2 is the i^{th} diagonal element of the matrix $\boldsymbol{\Sigma}$, and $\delta_0\left(\boldsymbol{\Lambda}_{ij}\right)$ is the Dirac delta function for $\boldsymbol{\Lambda}_{ij}$ at zero (i.e. a point mass function with all mass concentrated at zero).

The joint posterior of these parameters is intractable, but it is trivial to devise a Gibbs sampler that samples sequentially from conditional posteriors that are of simple form:

Factor-based sign restrictions (FSR) algorithm

1. Sample $\boldsymbol{\phi}_i$ for $i = 1, ..., n$ from

$$\boldsymbol{\phi}_i | \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \mathbf{f}, \mathbf{y} \sim N_k\left(\overline{\mathbf{V}}_i\left(\sum_{t=1}^{T} \sigma_i^{-2} \mathbf{x}_t' \widetilde{\mathbf{y}}_{it}\right), \overline{\mathbf{V}}_i\right), \quad (12)$$

   where $\widetilde{\mathbf{y}}_{it} = \mathbf{y}_{it} - \boldsymbol{\Lambda}_i \mathbf{f}_t$ and $\overline{\mathbf{V}}_i^{-1} = \left(\underline{\mathbf{V}}_i^{-1} + \sum_{t=1}^{T} \sigma_i^{-2} \mathbf{x}_t' \mathbf{x}_t\right)$.

2. Sample $\boldsymbol{\Lambda}_i$ for $i = 1, ..., n$ from

$$\boldsymbol{\Lambda}_i | \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \mathbf{f}, \mathbf{y} \sim MTN_{\mathbf{a} < vec(\boldsymbol{\Lambda}) < \mathbf{b}}\left(\overline{\mathbf{H}}_i\left(\sum_{t=1}^{T} \sigma_i^{-2} \mathbf{f}_t' \widehat{\mathbf{y}}_{it}\right), \overline{\mathbf{H}}_i\right), \quad (13)$$

   where $\widehat{\mathbf{y}}_{it} \equiv \boldsymbol{\varepsilon}_{it} = \mathbf{y}_{it} - \phi_i \mathbf{x}_t$, $\overline{\mathbf{H}}_i^{-1} = \left(\underline{\mathbf{H}_i}^{-1} + \sum_{t=1}^{T} \sigma_i^{-2} \mathbf{f}_t' \mathbf{f}_t\right)$, and $\underline{\mathbf{H}}_i = diag\left(\underline{h}_{i1}, ..., \underline{h}_{ir}\right)$. Here we define $MTN\left(\bullet\right)$ to be the multivariate truncated Normal distribution, and $\mathbf{a}, \mathbf{b}$

are the vectors indicating the truncation points, with ij^{th} element:

$$(\mathbf{a}_{ij}, \mathbf{b}_{ij}) = \begin{cases} (-\infty, 0) & \text{if } S_{ij} = -1, \\ (0, \infty) & \text{if } S_{ij} = 1, \\ (0, 0) & \text{if } S_{ij} = 0, \\ (-\infty, \infty) & \text{otherwise,} \end{cases} \tag{14}$$

for $i = 1, ..., n$, $j = 1, ..., r$.

3. Sample $\mathbf{f}_t$ for $t = 1, ..., T$ from

$$\mathbf{f}_t | \mathbf{\Lambda}, \mathbf{\Sigma}, \mathbf{\Phi}, \mathbf{y} \sim N\left(\overline{\mathbf{G}}\left(\mathbf{\Lambda}\mathbf{\Sigma}^{-1}\widehat{\mathbf{y}}_t\right), \overline{\mathbf{G}}\right), \tag{15}$$

where $\overline{\mathbf{G}}^{-1} = (\mathbf{I}_r + \mathbf{\Lambda}'\mathbf{\Sigma}\mathbf{\Lambda})$. Post-process the draws of the $T \times r$ matrix $\mathbf{f} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$ such that its r columns (corresponding to structural shocks) are uncorrelated and standardized to unit variance. This is done by applying first the Gram-Schmidt procedure and subsequently dividing each column of $\mathbf{f}$ with its standard deviation.

4. Sample σ_i^2 for $i = 1, ..., n$ from

$$\sigma_i^2 | \mathbf{\Lambda}, \mathbf{f}, \mathbf{\Phi}, \mathbf{y} \sim inv - Gamma\left(\frac{T}{2} + \underline{\rho}_i, \left[\underline{\kappa}_i^{-1} + \sum_{t=1}^{T} (\mathbf{y}_{it} - \phi_i \mathbf{x}_t - \mathbf{\Lambda}_i \mathbf{f}_t)' (\mathbf{y}_{it} - \phi_i \mathbf{x}_t - \mathbf{\Lambda}_i \mathbf{f}_t)\right]^{-1}\right) \tag{16}$$

A major aspect of the proposed Gibbs sampler algorithm for structural VAR inference is that all draws from the space of sign restricted reduced-rank matrices $\mathbf{\Lambda}$ are accepted.

2.2 Priors and efficient sampling

While the Gibbs sampler presented above is based on standard sampling steps for Bayesian VARs and static factor models (Lopes and West, 2004), I discuss here how to build an even more reliable algorithm that can handle larger models with the same ease it can estimate smaller models. The main computational concerns with the core algorithm presented above stem from the high dimensions of $\mathbf{\Phi}$[6] and the fact that $\mathbf{\Lambda}$ is both latent and has to be sampled from a restricted (truncated) Normal distribution.

[6]For example, in a 50 variable VAR with 4 lags and an intercept this matrix has 10,000+ elements.

First, overall the proposed factor-VAR had disturbances that have a diagonal covariance matrix. Therefore, the parameters $\mathbf{\Phi}$ can be sampled equation-by-equation (which is what Baumeister and Hamilton, 2015 also do in the SVAR setting) using (12). Instead of having to sample one large vector of nk parameters, n independent univariate autoregressions with k parameters each can be estimated – a step that is also trivial to parallelize using modern computers. This point has been established recently in Carriero et al. (2019).

Second, I additionally follow ideas in Bhattacharya et al. (2016) and use an efficient sampler for the Normal distribution, which makes full use of the Woodbury identity in order to sample from equation (12)efficiently. The idea is that during sampling from the Normal distribution one has to obtain the Cholesky of a large matrix $(\overline{\mathbf{V}}_i)$ which requires $\mathcal{O}(k^3)$ algorithmic operations. The transformation of Bhattacharya et al. (2016) allows sampling with worst case algorithmic complexity of $\mathcal{O}(T^2k)$ operations. Therefore, gains from replacing this step with the Bhattacharya et al. (2016) approach can be very substantial as either the dimension n of the VAR or the number of lags p increase. In particular, as $k = np+1$ becomes larger than T, a significantly smaller number of algorithmic operations will be required in order to sample from the posterior of $\mathbf{\Phi}$.

Third, as the dimension of $\mathbf{\Phi}$ increases polynomially in n and p, computation is not the sole concern; regularization also becomes an important aspect of statistical inference. Traditionally, empirical Bayes priors such as the Minnesota prior, have been used both in reduced-form and structural VARs; see Giannone et al. (2015) and Baumeister and Hamilton (2015), respectively. However, little theory exists on the high-dimensional shrinkage properties and posterior consistency of such empirical Bayes rules. In contrast, the *horseshoe prior for sparse signals* proposed by Carvalho et al. (2010) has been shown to lead to Bayes estimates that are consistent a-posteriori, and that attain a risk equivalent to the (Bayes) oracle; see Armagan et al. (2013) and Ghosh et al. (2016) and references therein.[7] Using equation (8), I

[7] van der Pas et al. (2014) also show that the horseshoe has good frequentist properties and can attain minimax-adaptive risk up to a constant, for squared error ($\updownarrow_2$) loss. Recently, Kowal et al. (2019) also establish the excellent shrinkage properties of the horseshoe in a dynamic linear model for time-series data.

specify the prior covariance matrix to have the following hierarchical structure

$$\begin{aligned}
\boldsymbol{\phi}_i|\sigma_i^2\tau_i^2\boldsymbol{\Psi}_i &\sim N_k\left(\mathbf{0}, \underline{\mathbf{V}}_i\right), && (17)\\
\underline{\mathbf{V}}_{i,(jj)} &= \sigma_i^2\tau_i^2\psi_{i,j}^2, && (18)\\
\psi_{i,j} &\sim Cauchy^{+}\left(0,1\right), && (19)\\
\tau_i &\sim Cauchy^{+}\left(0,1\right), && (20)
\end{aligned}$$

where $\underline{\mathbf{V}}_{i,(jj)}$ denotes the j^{th} diagonal element of prior covariance matrix $\underline{\mathbf{V}}_i$, $\boldsymbol{\Psi}_i = diag\left(\psi_{i,1}^2, ..., \psi_{i,k}^2\right)$, and $Cauchy^{+}$ denotes the half-Cauchy distribution with support on the set of positive real numbers $\mathbb{R}^{+}$.[8] The most important aspect of the horseshoe prior is that it requires absolutely no input from the researcher, while retaining at the same time its excellent shrinkage properties. There are various reparametrizations of this prior and associated MCMC sampling schemes. Here I follow Neal (2003) and use a slice sampler (within the Gibbs sampler algorithm) that allows to update $\psi_{i,(jj)}$ and τ_i efficiently in high-dimensions.

Fourth, a major challenge in the proposed algorithm is that in Step 2, $\boldsymbol{\Lambda}_{ij}$ has to be sampled from a multivariate truncated Normal distribution which is notoriously hard to simulate from (Geweke, 1996). In order to deal with this significant computational challenge, I follow Geweke (1996) and sample each $\boldsymbol{\Lambda}_{ij}$ conditional on all remaining $\boldsymbol{\Lambda}_{-ij}$ elements. Most importantly, I adopt the exact minimax tilting method for generating i.i.d. data from a truncated normal distribution that was recently proposed by Botev (2017). This method allows orders of magnitude computational improvements relative to the truncated normal sampler proposed by Geweke (1996) and others; see Botev (2017) for a review of this literature. Given absence of prior information on the variance of $\boldsymbol{\Lambda}$ (above and beyond the information we have on the sign restrictions) I set $\underline{h}_{ij} = 4$ in the remainder of the paper,

[8]Note that the prior covariance matrix of $\boldsymbol{\phi}_i$ is a function of the VAR variance σ_i^2. This is done in order to enhance numerical stability when the endogenous variables in the VAR are measured in different units, even though in practice it is trivial to specify this prior to be independent of σ_i^2 (which would be essential in case we want to specify σ_i^2 to be time-varying using, for instance, a stochastic volatility specification.

which is a noninformative choice for the parameters of a loadings matrix.

All the considerations above ensure that the algorithm proposed in the previous subsection is not only fast, but is also numerically stable and can scale up to much larger VAR dimensions than ever considered before in the literature. Note that there are remaining prior settings and algorithmic steps (those involving parameters $\mathbf{f}_t$ and $\mathbf{\Sigma}$), but these steps are already computationally trivial and may not be made more efficient. Additionally, the prior for $\mathbf{f}_t$ is fixed by the required identification restrictions, and the $\mathbf{\Sigma}$ are integrated out with fairly noninformative values (I set $\rho_i = 1$ and $\kappa_i = 0.1$ for all i, in all VARs estimated in this paper). Exact details of the proposed algorithm, including the additional enhancements described in this subsection, are provided in the online Appendix.

2.3 Likelihood-based testing of identifying assumptions in VARs

Our modeling assumption that the structural shocks $\mathbf{f}_t$ are latent parameters that have to be estimated from the likelihood, has an enormous implication: sign restrictions become part of the model likelihood and can be explicitly tested the same way economists test other parametric restrictions in regression models (e.g. inequality constraints as in Geweke, 1996). This concept might not make sense for economic shocks that are indisputable such as the effects of an aggregate demand/supply shock, however, there are cases of shocks where the expected sign might not be known a-priori with certainty. Alternatively, a researcher might want to statistically test the plausibility of certain zero restrictions, or simply compare the performance of two different VAR models (e.g. with different number of lags and/or variables) given the same set of identifying restrictions. In the context of the proposed VAR model with factor structure, all these cases can be tested explicitly using marginal likelihoods or Bayes factors. Parametric (i.e. zero or sign) restrictions on $\mathbf{\Lambda}$ that agree with the information in the data will result in higher marginal data likelihood relative to restrictions that are not supported by the data. Put differently, given the decomposition in equation (3), plausible restrictions in $\mathbf{\Lambda}$ will result in estimation of a more precise unconditional VAR covariance

matrix $\mathbf{\Omega}$.

Marginal likelihoods are not numerically stable in high-dimensional VARs (Giannone et al., 2015) and they can be demanding to compute even in smaller VARs with many layers of latent parameters (e.g. hierarchical priors like the horseshoe; stochastic variances; Markov-switching coefficients). In order to deal with this computational aspect, I instead propose to calculate the Deviance Information Criterion (DIC) of Spiegelhalter et al. (2002) as a default criterion for comparing parametric restrictions on $\mathbf{\Lambda}$. For the matrix of VAR model parameters $\mathbf{\Theta} = \left(\mathbf{\Phi}, \mathbf{\Lambda}, \{\mathbf{f}_t\}_{t=1}^{T}, \mathbf{\Lambda}\right)$ the DIC is defined as the quantity

$$DIC = -4E_{p(\mathbf{\Theta}|\mathbf{y})}\left(\log f\left(\mathbf{y}|\mathbf{\Theta}\right)\right) + 2\log f\left(\mathbf{y}|\widehat{\mathbf{\Theta}}\right), \tag{21}$$

where $\log f\left(\mathbf{y}|\mathbf{\Theta}\right)$ is the log of the likelihood function implied by the regression in (7). The first term in the criterion above is the expectation of the likelihood w.r.t the parameter posterior, and can be obtained numerically using Monte Carlo integration by simply evaluating the likelihood at each MCMC draw from the posterior of the parameters $\mathbf{\Theta}$. The second term is the likelihood function evaluated at an estimate $\widehat{\mathbf{\Theta}}$ of high posterior density (typically posterior mean or mode). As with all other information criteria used in statistics/econometrics, lower values signify better fit. The DIC is not a first-order approximation to the marginal likelihood, in the same way that the Bayesian Information Criterion (BIC) is. The marginal likelihood, also known as *prior predictive distribution*, addresses the issue of how well the data are predicted by the priors. In this sense, the DIC is a criterion that is closely related to measuring fit according to the *posterior predictive distribution*, rather than marginal likelihoods. As a result, for the purpose of assessing the fit of a VAR that is intended to be used for out-of-sample projections (impulse responses, forecast error variance decompositions etc), the DIC can be considered as a more appropriate predictive measure of fit compared to marginal likelihoods or alternative in-sample measures of fit.

3 Simulation study

In this section, the properties of the new algorithm are explored using artificially generated data. The core exercise involves generating multivariate time series from a data generating process (DGP) that fully matches equation (7), and estimating parameters and impulse response functions based on time series generated from this DGP. I first implement this experiment assuming that a correctly specified model is estimated using the artificial data. Subsequently, various cases of misspecification errors during the estimation process are considered – that is, I estimate models that do not perfectly match the correct DGP.

The DGP is of the form

$$\mathbf{y}_t = \widehat{\mathbf{\Phi}}\mathbf{x}_t + \widehat{\mathbf{\Lambda}}\mathbf{f}_t + \mathbf{v}_t, \text{ for } t = 1, ..., \widehat{T}, \tag{22}$$

$$\mathbf{v}_t \sim N\left(\mathbf{0}, \widehat{\mathbf{\Sigma}}\right), \quad \mathbf{f}_t \sim N\left(\mathbf{0}, \mathbf{I}\right), \tag{23}$$

$$\mathbf{y}_{(-p+1):0} = \mathbf{0}, \quad p = 12, \quad r = 3. \tag{24}$$

The DGP parameters $\widehat{\mathbf{\Phi}}$, $\widehat{\mathbf{\Lambda}}$, $\widehat{\mathbf{\Sigma}}$ are based on estimates of a VAR on real data. First, monthly data on 14 monthly macroeconomic variables are collected[9] for the US over the period 1965M1 - 2007M12, providing $\widehat{T} = 516$ observations.[10] At a second step, an estimate $\widehat{\mathbf{\Phi}}$ is obtained by applying OLS to an unrestricted VAR(12) estimated with these 14 observed US variables. The third step is to obtain the first $r = 3$ principal components of these OLS residuals, and store the estimate $\widehat{\mathbf{\Lambda}}$ using OLS in a regression between the VAR residuals and their principal components. Finally, the residuals from this latter regression provide the elements of the diagonal matrix $\widehat{\mathbf{\Sigma}}$, by means of equation-by-equation application of the usual least squares formula for the variance.

While it is not possible, or even interesting, to print all estimates $\widehat{\mathbf{\Phi}}$ used as input in

[9]The variables are: 1) real GDP, 2) GDP deflator, 3) federal funds rate, 4) commodity price index, 5) total reserves, 6) nonborrowed reserves, 7) S&P 500, 8) M1 , 9) unemployment rate , 10) industrial production , 11) employment , 12) CPI , 13) core CPI , 14) core PCE. More details on these variables is provided in the online Appendix.

[10]In practice, I generate $\widehat{T} + 1000$ observations and discard the first 1000 observations.

the DGP, it is instead interesting to look at the estimates $\widehat{\mathbf{\Lambda}}$ obtained using the procedure described above. This is because both the signs and the magnitudes of the implied IRFs in the true DGP will be affected by those estimates. Panel (A) of Table 1 shows the OLS estimates, where the diagonal is normalized to be one, by dividing each element in the m^{th} column of $\widehat{\mathbf{\Lambda}}$ with the original value of its m^{th} element, $m = 1, 2, 3$. While this matrix is the outcome of using real data and applying simple principal components plus OLS estimation (which carry no economic restrictions), the signs implied by it allow us to classify the three pseudo-shocks as aggregate supply, aggregate demand, and monetary policy, respectively. The estimated magnitudes of course are not necessarily economically meaningful. Nevertheless, this is an exercise where the main aim is to check the numerical precision of the new algorithm, so the estimates in panel (A) of Table 1 are perfectly valid inputs for a DGP. Finally, Panel (B) of Table 1 shows the sign restrictions imposed on $\mathbf{\Lambda}$. These comply with the signs imposed in the DGP, and in 11 instances no sign restrictions are imposed (these entries are denoted as *NA*).

	(A) True parameter values			(B) Sign restrictions		
Variable	1^{st} shock	2^{nd} shock	3^{rd} shock	1^{st} shock	2^{nd} shock	3^{rd} shock
real GDP growth	1.00	-1.39	-0.87	+	–	–
GDP deflator inflation	1.42	1.00	-0.71	+	+	–
Fed funds rate	0.49	-0.28	1.00	NA	NA	+
Commodity prices	0.16	0.16	-0.45	NA	NA	–
Total reserves	-0.61	0.22	-3.48	NA	NA	NA
Nonborrowed reserves	-0.91	0.25	-3.37	NA	NA	–
Stock prices	-0.25	-0.30	-0.82	NA	NA	–
M1	-1.03	-0.48	-1.27	–	–	–
Unemployment	-0.63	0.51	0.43	–	+	+
Industrial production	1.12	-1.34	-0.87	+	–	–
Employment	0.88	-1.00	-1.01	+	–	–
CPI inflation (total)	1.44	1.01	-0.75	+	+	–
CPI inflation (core)	1.05	0.49	-1.12	+	+	–
PCE inflation (core)	1.05	0.57	-0.80	+	+	–

Notes: Panel (A) shows true parameter values used as input in the data generating process (DGP), while panel (B) shows the sign restrictions imposed during econometric estimation using each artificial dataset from the DGP. Entries in panel (B) show the restrictions imposed: + for positive sign; – for negative sign; NA for no restriction.

Table 1: *OLS estimates* $\widehat{\mathbf{\Lambda}}$ *used in the DGP, and sign restrictions used for estimation*

For estimation purposes five different scenarios are assumed: one correctly specified case and four misspecified cases. These are denoted as C1-C5, and are defined as follows:

C1 Correctly specified model with $n = 14$ dependent variables, $p = 12$ lags, $r = 3$ shocks.

C2 Misspecified model with $n = 8$, using the first eight variables in Table B1 in the online Appendix. All other settings are correct, that is, $p = 12$ lags, $r = 3$ shocks.

C3 Misspecified model with $p = 2$ lags. All other settings are correct, that is, $n = 14$ and $r = 3$.

C4 Misspecified model with $r = 2$ shocks, using only the restrictions on the first two shocks in panel (B) of Table 1. All other settings are correct, that is, $n = 14$ and $p = 12$.

C5 Misspecified model with $r = 4$ shocks, using an additional shock.[11] All other settings

[11] This fourth shock is identified using the randomly selected vector of restrictions $s = [+, +, +, -, +, NA, NA, -, +, +, +, +, +, +]$.

are correct, that is, $n = 14$ and $p = 12$.

500 datasets of size $T = 516$ are generated and posterior mean estimates of all parameters, IRFs and DICs from all five cases above are obtained. Results presented next are based on the distribution of the posterior means over these 500 artificial datasets.

Before evaluating precision of estimates over the Monte Carlo iterations, it is important to first evaluate general model fit using the DIC. Table 2 shows the value of the deviance information criterion attained by each of the five cases. Because case C2 refers to a VAR with $n = 8$, it is impossible to directly compare it with the other four cases that assume $n = 14$.[12] For that reason I present two DIC metrics, a full one based on all $n = 14$ variables (with no value available for C2) and a reduced DIC which is the same formula evaluated only on the first eight VAR equations (which are common to all five cases). These are labelled in Table 2 as DIC_{14} and DIC_8, respectively. According to both subsets of criteria, the correctly specified estimated model case C1 is the best one as it attains the lowest DIC value. Interestingly, the case where an additional fourth shock is incorrectly estimated (C5), doesn't seem to harm estimation accuracy; at least not as much as the case of estimating one less shock (C4). By far the worst type of misspecification seems to be the one related to the lag-length. This is a characteristic of the VAR model rather than a "problem" with the specific algorithm or prior. As long as the true DGP has $p = 12$ important lags, estimating the VAR with $p = 2$ provides a huge loss of information. In contrast, reducing the VAR from $n = 14$ (which is the truth in the DGP) to $n = 8$ as in case C2, harms much less the fit of the VAR.

[12] Information criteria can only be used to compare models with the same dependent variable $\mathbf{y}$.

	C1	C2	C3	C4	C5
DIC_{14} value	7393.14	n/a	27241.13	11859.45	9037.44
DIC_8 value	15300.63	17094.28	28630.21	27981.48	19269.55

Notes: DIC_{14} is the deviance information criterion applied jointly to all 14 VAR equations. DIC_8 is the same criterion applied jointly only to the first 8 VAR equations. Case C2 does not have a DIC_{14} value because it assumes that the VAR has $n = 8$ variables.

Table 2: *DIC values attained by the correctly specified and misspecified models estimated on artificial data*

Next, estimation accuracy of the proposed algorithm has to be evaluated. Since the main focus of sign restrictions algorithms is on impulse response analysis, I compare precision of the estimated impulse response functions using all generated datasets. IRFs are combinations of all VAR parameters $\mathbf{\Phi}, \mathbf{\Lambda}, \mathbf{f}, \mathbf{\Sigma}$, therefore comparing their precision provides a convenient summary of overall estimation precision in a VAR model. Figure 1 shows the responses of the first three variables in the VAR to the three identified pseudo-shocks, in the correctly specified case (C1). Green solid lines are medians over the posterior IRFs in the 500 estimated VARs using an equal number of artificial datasets. Shaded areas show the 90% probability bands of these IRFs. Finally, black dashed lines show the true IRFs implied by the parameters that are fed into the DGP. The 90% bands always include the true IRF, which suggests that estimation precision is satisfactory. The online Appendix shows identical graphs for the four misspecified cases C2-C5. These graphs become a visual confirmation of the numerical results in Table 2, that is, case C5 quite precisely captures the path of the true IRFs, while case C3 results in the largest estimation errors.

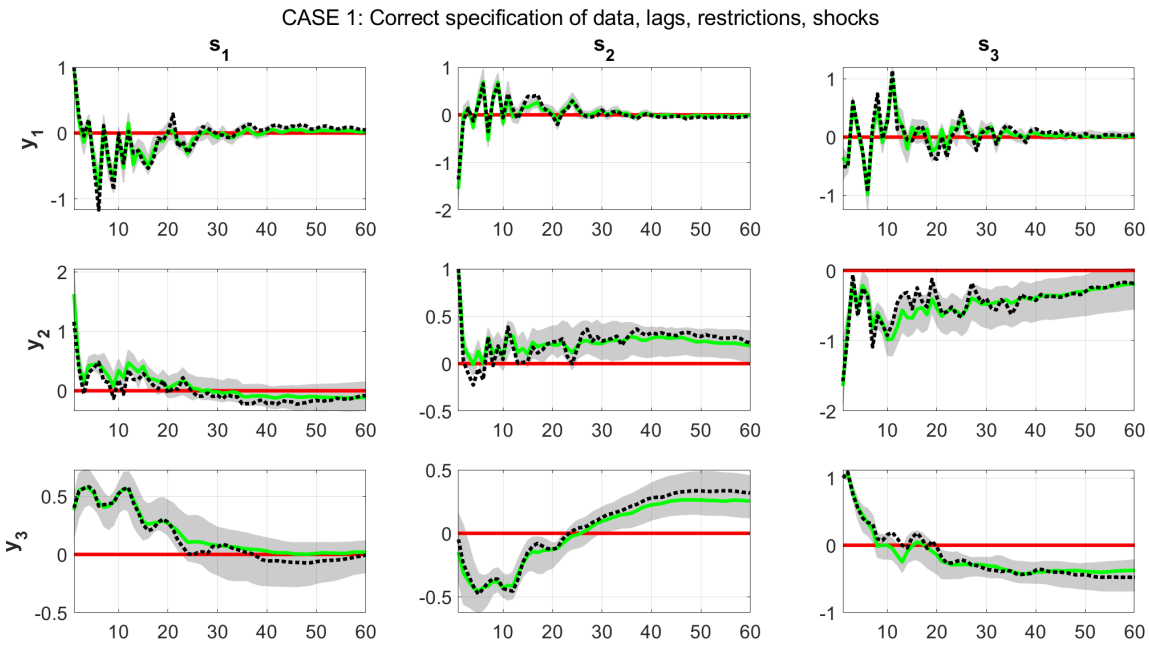


Figure 1: *Impulse response functions of the first three artificially generated variables (denoted as y_1, y_2, y_3) in response to the three identified shocks (denoted as s_1, s_2, s_3) in model C1 (correctly specified model). The green solid lines show the posterior median IRFs over the 500 Monte Carlo iterations, and the gray shaded areas their associated 90% bands. The true IRFs based on the DGP are shown using the black dashed lines.*

3.1 How fast is the new algorithm?

The next Section makes clear that in the context of the empirical application in Furlanetto et al. (forthcoming), the new algorithm is multiple times faster than the algorithm of Rubio-Ramírez et al. (2010) in a six-variable VAR with five identified shocks. Nevertheless, it would be interesting to use artificial data in order to provide more thorough evidence on how fast the factor sign restrictions algorithm is, and how large a VAR it can scale to. For that reason artificial data are generated from the same DGP described in equations (22)-(23) for various values of the key parameters that affect the dimensionality of the VAR, namely T, n and r. Due to the fact that this exercise pushes the VAR dimension n to very large values, I fix $p = 1$ in order to be able to ensure that the VAR process in the DGP is always stationary, and generation of explosive data is excluded. For the purposes of this exercise I set $\mathbf{\Phi} = 0.9\mathbf{I}_n$,

$\mathbf{\Lambda}_{ij} \sim U(-1,1)$ and $\Sigma_i \sim U(0,1)$, for all $i = 1,...,n$ and $j = 1,...,r$. During estimation nk sign restrictions are imposed, simply by obtaining the signs of the randomly generated matrix $\mathbf{\Lambda}$.[13]

	$T = 200$			$T = 500$		
	$n = 15$	$n = 50$	$n = 100$	$n = 15$	$n = 50$	$n = 100$
$r = 3$	1	2	6	4	11	23
$r = 10$	1	4	8	4	12	23
$r = 20$	NA	6	11	NA	15	25

Table 3: *Computer time in minutes (defined as (seconds/60), rounded to the nearest integer) for obtaining* $10,000$ *post-burn-in draws (*$12,000$ *in total) using various VAR sizes. Here* T *is the number of observations,* n *the number of endogenous variables, and* r *the number of shocks. All VARs have* $p = 1$ *lag.*

Table 3 shows the average, over 10 Monte Carlo iterations, machine time in minutes (defined as the total estimation time in seconds divided by 60 and then rounded to the nearest integer) needed to obtain $10,000$ draws from the posterior of all parameters after discarding $2,000$ draws (hence, $12,000$ draws in total). These results show that in a huge-dimensional VAR with $n = 100$ series, $T = 500$ observations, and $r = 20$ shocks, it only takes 25 minutes to obtain $10,000$ draws from all parameter matrices, including the 1000 sign-restricted elements in $\mathbf{\Lambda}$. For the smaller model with $n = 15$ – which is already much larger than the vast majority of models considered in the sign restrictions literature – it only takes less than five minutes to obtain the same number of draws when $T = 500$, and only one minute when $T = 200$. These fantastic timings justify the choice to focus on carefully developing a Gibbs sampler that is computationally efficient.[14]

[13]The purpose of this exercise is not to estimate meaningful restrictions, rather just to measure times. In this case, I impose the maximum number of restrictions possible on $\mathbf{\Lambda}$ in order to test the new algorithm in a worst-case scenario where all nk of its elements are restricted and have to be generated from a truncated Normal posterior.

[14]The Gibbs sampler typically loses efficiency when there is high correlation in the samples from the posterior. In the online Appendix I show that, in order to draw $\mathbf{\Lambda}_{ij}$ from univariate (instead of the intractable multivariate) truncated Normal conditionals, we need to condition on $\mathbf{\Lambda}_{-ij}$, i.e. the set of all elements of $\mathbf{\Lambda}$ excluding the ij^{th}. This conditioning increases correlation relative to sampling directly the full matrix $\mathbf{\Lambda}$. However, inefficiency factors for the Gibbs sampler in the linear factor-VAR specification are still quite low

The results above are based on code written in MATLAB2019b and run in a personal computer with Intel Core i7 8700K, tuned at 4.9Ghz, and 32GB of RAM. Note that the Gibbs sampler algorithm iterates over each VAR equation independently and, thus, significant speed improvements can be achieved by taking advantage of parallel processing abilities of modern computers and high-performance clusters (HPCs). In MATLAB this is as simple as replacing *for loops* with *parfor loops*. Therefore, the algorithm indeed allows the estimation of arbitrarily large VAR models, as it is claimed in the Introduction.

In practical situations, the only issue that might inhibit the performance of the algorithm (and any Monte Carlo-based algorithm, to that effect) is the fact that in very large dimensions we may be sampling parameters $\mathbf{\Phi}$ in a region of the posterior that implies nonstationarity of the VAR. In order to make sense out of impulse response functions, forecast error variance decompositions, historical decompositions etc, we need to make sure we maintain only samples from the posterior which are stationary. For that reason it is important to stress that, throughout my experiments, the horseshoe prior does a great job (especially relative to a subjectively chosen Minnesota prior) in shrinking the coefficients $\mathbf{\Phi}$ towards a more numerically stable region of their posterior, where the VAR model is stationary.

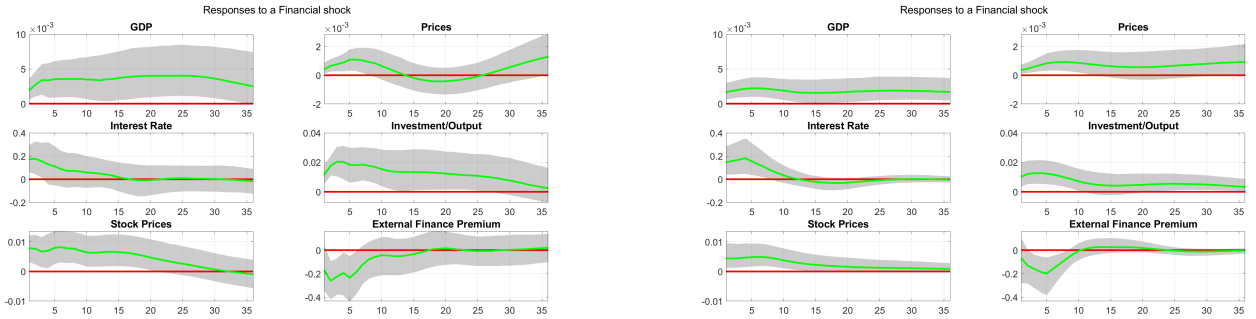
4 A (reasonably) large-scale VAR model for measuring financial shocks

In this section I revisit the empirical exercise in Furlanetto et al. (forthcoming), who aim to measure various financial shocks to the US economy.[15] Given computational restrictions,

(MCMC diagnostic results are available upon request). Additionally, given the ability of the algorithm to obtain quickly tens of thousands of draws from the posterior, concerns about possible correlation of draws can be alleviated by doing "thinning" – i.e. the procedure of storing only every ρ^{th} sample from the posterior, where is ρ is the order of the highest significant autocorrelation in the chain.

[15]The online Appendix provides the results of an additional numerical exercise (*measuring optimism shocks*) that builds on Arias et al. (2018).

due to their use of the Rubio-Ramírez et al. (2010) accept/reject algorithm, Furlanetto et al. (forthcoming) end up estimating a series of smaller VARs in order to sequentially measure and label interesting financial shocks, such as uncertainty, credit and housing. Before illustrating how to use the new algorithm to collectively measure all these shocks in one high-dimensional data setting, we first try to replicate their benchmark results. Among all VAR specifications, Furlanetto et al. (forthcoming) specify a *baseline* VAR specification with $p = 5$ lags, using data on real GDP, consumer prices, interest rate, investment-to-output ratio, stock prices, and the external finance premium.[16] All data are for the 1985Q1 - 2013Q2 period. The online Appendix provides exact details of all series and transformations used, which in this case they are identical to those reported in Furlanetto et al. (forthcoming, Table 11).



(a) Rubio-Ramírez et al. (2010) algorithm

(b) Factor-based sign restrictions algorithm

Figure 2: *This figure replicates the impulse response functions to a financial shock using the baseline specification of Furlanetto et al. (forthcoming). Panel (a) shows results based on the exact configuration of Furlanetto et al. (forthcoming, see Figure 1), using the algorithm of Rubio-Ramírez et al. (2010). Panel (b) replicates the same financial shock using the new sign restrictions algorithm.*

Figure 2 shows the effects of a financial shock identified as a shock that causes GDP, consumer prices, stock prices, interest rate and the investment/output ratio to react positively contemporaneously. The sign of the spread is left unrestricted. Panel (a) replicates the impulse responses also shown in Figure 1 of Furlanetto et al. (forthcoming). Panel (b) shows the same

[16]The external finance premium is defined as the spread between yields on Baa rated bonds and the federal funds rate. Notice that the three variables that are not already expressed as rate, ratio, or spread (i.e. GDP, consumer prices, and stock prices), are transformed only using logarithms of the levels and not growth rates – see the online Appendix for exact definitions, transformations, and data sources. Also note that these authors use a noninformative (uniform) prior, while I use the shrinkage horseshoe prior described in Section 2.

responses produced by application of the new algorithm for sign restrictions. While there are obvious differences in how wide the bands of the IRFs are (due to the underlying assumptions about the VAR likelihood and the prior distributions), the results are qualitatively very similar and assuring that the new algorithm produces sensible results. Following up on the discussion in the previous section, it takes roughly four hours to obtain 2000 draws from the Furlanetto et al. (forthcoming) using their MATLAB code and exact numerical settings based on the Rubio-Ramírez et al. (2010) algorithm. In contrast, it takes less than five minutes to obtain $60,000$ draws from the proposed Gibbs sampler (where out of these $60,000$ draws we discard $10,000$ and then save every 10^{th} draw, leading to $5,000$ draws from the posterior of VAR parameters and impulse response functions).

We next proceed to demonstrate how the new algorithm can estimate one, large-dimensional system in order to measure in one setting all the financial shocks that Furlanetto et al. (forthcoming) identify. The larger VAR that these authors specify has seven variables and six shocks: aggregate supply, aggregate demand, investment, housing, uncertainty, and credit. These authors do not identify a monetary shock using this larger VAR, possibly due to computational concerns. Here we attempt to use all available variables in Furlanetto et al. (forthcoming) to identify seven shocks, that is, the six shocks just listed plus a monetary shock. We also use additional measures of output, consumer prices, stock prices, interest rate, and credit spread, in order to enhance identification. We end up with a 15-variable VAR with $p = 5$ on the following variables: 1) real GDP; 2) prices (GDP deflator); 3) interest rate (3-month Tbill); 4) investment to output ratio; 5) stock prices (real S&P500 prices); 6) credit spread (Baa minus Fed funds rate); 7) credit to real estate value ratio; 8) excess bond premium (EBP); 9) EBP to VIX ratio; 10) mortgage rate (30-year rates); 11) employment; 12) Federal funds rate; 13) core CPI; 14) stock prices 2 (real DJIA prices); and 15) credit spread 2 ("GZ" spread). The online Appendix has detailed definitions of these variables, transformations used, and sources.

Table 4 shows the signs imposed on each of the 15 variables in order to identify each of the

seven structural shocks. This is a large matrix of restrictions, but the new algorithm can handle computationally the task of drawing $60,000$ samples from the posterior of all parameters (including the structural matrix of contemporaneous shocks) in a matter of minutes. As it was the case with the baseline VAR above, out of these $60,000$ draws $10,000$ are discarded and every 10^{th} sample is stored, resulting in $5,000$ samples used to produce numerical results from this large model. The horseshoe prior also has a crucial role in the estimation of this model, as we have 1140 parameters in $\mathbf{\Phi}$ and only 114 observations for each of the 15 endogenous variables.

	SHOCKS						
	Supply	Demand	Monetary	Investment	Housing	Uncertainty	Credit
ORIGINAL VARIABLES IN FURLANETTO ET AL. (FORTHCOMING):							
GDP	+	+	+	+	+	+	+
prices	–	+	+	+	+	+	+
interest rate	NA	+	–	+	+	+	+
investment/output	NA	–	NA	+	+	+	+
stock prices	+	NA	NA	–	+	+	+
spread	NA	NA	NA	NA	NA	NA	NA
credit/real estate value	NA	NA	NA	NA	–	+	+
EBP	NA	NA	NA	NA	NA	–	–
EBP/VIX	NA	NA	NA	NA	NA	+	–
mortgage rates	NA	NA	NA	NA	NA	NA	–
ADDITIONAL MEASURES OF OUTPUT, PRICES ETC.:							
employment	+	+	+	+	+	+	+
Federal funds rate	NA	+	–	+	+	+	+
core prices	–	+	+	+	+	+	+
stock prices 2	+	NA	NA	–	+	+	+
spread 2	NA	NA	NA	NA	NA	NA	NA

Notes: Entries in this table show the restrictions imposed: + positive sign; – negative sign; NA no restriction.

Table 4: *Identified shocks and sign restrictions imposed on the matrix* $\mathbf{\Lambda}$ *in the large, 15-variable VAR model*

Figure 3 shows the impulse responses of the 15 endogenous variables to a credit shock. The green lines are posterior medians, and the shaded areas 68% bands. The magnitudes and shapes of the IRFs are consistent with the ones reported in Furlanetto et al. (forthcoming, Figure 7), despite the fact that in the case of variables such as GDP the IRFs are strongly different from zero. The most interesting feature of this figure is the effect of a credit shock

on the two credit spread variables we used in the same VAR. Furlanetto et al. (forthcoming) use these spreads (plus an additional third spread we haven't included here) one at a time in their VAR in order to assess robustness of their results. These authors do not impose sign restrictions on the credit spread and they find that in their baseline specification this tends to be negative. In the large VAR case, the first credit spread variable has a strong negative contemporaneous response before subsequently moving to positive territory, while the second credit variable does not have a contemporaneous response different from zero and in subsequent period reacts positively. Such results show the important avenues for identifying various structural shocks that the new algorithm opens up: by using large information sets we can have the ability to identify several structural shocks in one setting, thus, making comparisons and testing of structural hypotheses more transparent. The online Appendix provides figures of the impulse responses to the remaining identified six shocks in the large, 15-variable VAR setting.

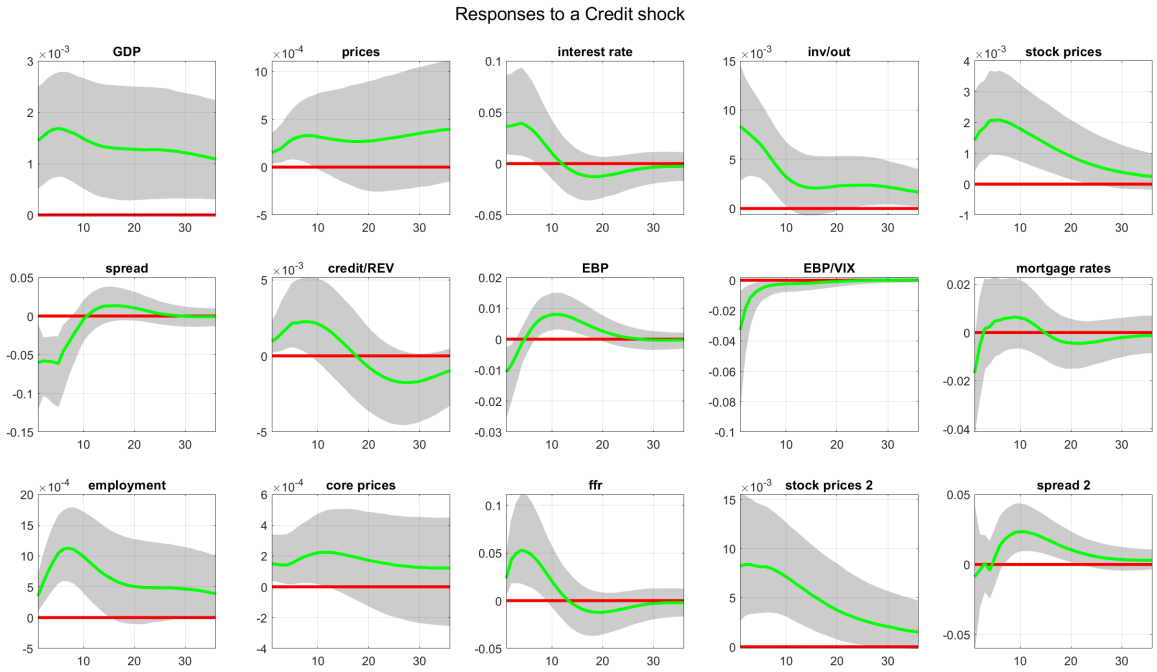


Figure 3: *Impulses response functions to a credit shock in the large, 15-variable VAR with seven shocks identified in total.*

5 Conclusions

This paper outlines a new algorithm based on a VAR methodology that fully utilizes the interpretability and parsimony of factor models. In particular, the novel element of the proposed approach is the formulation of reduced-form VAR disturbances using a common factor structure, and the derivation of an algorithm that allows for efficient sampling of sign-restricted decompositions of the VAR covariance matrix. The new algorithm can handle VARs with possibly 100 or more variables and it provides sensible numerical results compared to the algorithm of Rubio-Ramírez et al. (2010) – despite the fact that the two algorithms rely on different modeling assumptions and are not directly comparable.[17] Therefore, the new algorithm can be seen as a useful tool in the toolbox of modern macroeconomists, that complements existing algorithms and that opens up new avenues for empirical research using large-scale VAR models.

References

Ahmadi, P. A. and H. Uhlig (2015): "Sign Restrictions in Bayesian FaVARs with an Application to Monetary Policy Shocks," Working Paper 21738, National Bureau of Economic Research.

Anderson, T. W. and H. Rubin (1956): "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, Berkeley, Calif.: University of California Press, 111–150.

Arias, J. E., D. Caldara, and J. F. Rubio-Ramírez (2019): "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure," *Journal of Monetary Economics*, 101, 1 – 13.

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018): "Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications," *Econometrica*, 86, 685–720.

[17] Additional numerical results are provided in the online Appendix.

Armagan, A., D. B. Dunson, J. Lee, W. U. Bajwa, and N. Strawn (2013): "Posterior Consistency in Linear Models Under Shrinkage Priors," *Biometrika*, 100, 1011–1018.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

Baumeister, C. and J. D. Hamilton (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963–1999.

Beaudry, P., D. Nam, and J. Wang (2011): "Do Mood Swings Drive Business Cycles and is it Rational?" Working Paper 17651, National Bureau of Economic Research.

Bernanke, B. S. and A. S. Blinder (1992): "The Federal Funds Rate and the Channels of Monetary Transmission," *The American Economic Review*, 82, 901–921.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach*," *The Quarterly Journal of Economics*, 120, 387–422.

Bhattacharya, A., A. Chakraborty, and B. K. Mallick (2016): "Fast Sampling with Gaussian Scale Mixture Priors in High-Dimensional Regression," *Biometrika*, 103, 985–991.

Botev, Z. I. (2017): "The Normal Law Under linear Restrictions: Simulation and Estimation via Minimax Tilting," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 125–148.

Canova, F. and M. Paustian (2011): "Business cycle measurement with some theory," *Journal of Monetary Economics*, 58, 345 – 361.

Carriero, A., T. E. Clark, and M. Marcellino (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137 – 154, big Data in Dynamic Predictive Econometric Modeling.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010): "The Horseshoe Estimator for Sparse Signals," *Biometrika*, 97, 465–480.

Furlanetto, F., F. Ravazzolo, and S. Sarferaz (forthcoming): "Identification of Financial Factors in Economic Fluctuations," *The Economic Journal*, Eearly View, doi:10.1111/ecoj.12520.

Geweke, J. F. (1996): "Bayesian Inference for Linear Models Subject to Linear Inequality Constraints," in *Modelling and Prediction Honoring Seymour Geisser*, ed. by J. C. Lee, W. O. Johnson, and A. Zellner, New York, NY: Springer New York, 248–263.

Ghosh, P., X. Tang, M. Ghosh, and A. Chakrabarti (2016): "Asymptotic Properties of Bayes Risk of a General Class of Shrinkage Priors in Multiple Hypothesis Testing Under Sparsity," *Bayesian Anal.*, 11, 753–796.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97, 436–451.

Gorodnichenko, Y. (2005): "Reduced-Rank Identification of Structural Shocks in VARs," SSRN working paper 590906, University of California, Berkeley.

Kilian, L. and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2019): "Dynamic Shrinkage Processes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81, 781–804.

Lopes, H. F. and M. West (2004): "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14, 41–67.

Matthes, C. and F. Schwartzman (2019): "What Do Sectoral Dynamics Tell Us About the Origins of Business Cycles?" Working Paper 19-9, Federal Reserve Bank of Richmond.

Mountford, A. and H. Uhlig (2009): "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics*, 24, 960–992.

Neal, R. M. (2003): "Slice sampling," *The Annals of Statistics*, 31, 705–767.

Ouliaris, S. and A. Pagan (2016): "A Method for Working with Sign Restrictions in Structural Equation Modelling," *Oxford Bulletin of Economics and Statistics*, 78, 605–622.

Ramey, V. (2016): "Chapter 2 - Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 71 – 162.

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

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002): "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.

Stock, J. H. and M. W. Watson (2005a): "Implications of Dynamic Factor Models for VAR Analysis," Working Paper 11467, National Bureau of Economic Research.

——— (2005b): "Understanding Changes in International Business Cycle Dynamics," *Journal of the European Economic Association*, 3, 968–1006.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381 – 419.

——— (2017): *Shocks, Sign Restrictions, and Identification*, Cambridge University Press, vol. 2 of *Econometric Society Monographs*, 95–127.

van der Pas, S. L., B. J. K. Kleijn, and A. W. van der Vaart (2014): "The horseshoe estimator: Posterior concentration around nearly black vectors," *Electronic Journal of Statistics*, 8, 2585–2618.

von Neumann, J. (1951): "Various Techniques Used in Connection with Random Digits," in *Monte Carlo Method*, ed. by A. S. Householder, G. E. Forsythe, and H. H. Germond, Washington, DC: US Government Printing Office, vol. 12 of *National Bureau of Standards Applied Mathematics Series*, chap. 13, 36–38.

Online Appendix to "Sign restrictions in high-dimensional vector autoregressions"

Dimitris Korobilis

University of Glasgow

A Full Gibbs sampler for VARs with sign restrictions

The Gibbs sampler presented in the main paper is a quite accurate description of the core algorithmic steps required in order to estimate the VAR model with factor structure in the residuals. Nevertheless, in order to produce empirical and other results, we have relied on the hierarchical horseshoe prior of Carvalho et al. (2010), two fast algorithms from drawing from the Normal (Bhattacharya et al., 2016) and truncated Normal (Botev, 2017) distributions, respectively, and the slice sampler of Neal (2003) in order to update the horseshoe prior parameters. Therefore, it is important to rewrite the Gibbs sampling algorithm in full, and give further explanations about the three enhancements that guarantee a fast and reliable algorithm in high dimensions.

We repeat the full prior specification, which now includes the hierarchical horseshoe prior

on $\boldsymbol{\phi}_i$. The priors for the i^{th} VAR equation, $i = 1, ..., n$ is:

$$\boldsymbol{\phi}_i|\sigma_i^2, \tau_i^2, \boldsymbol{\Psi}_i^2 \sim N_k\left(\mathbf{0}, \sigma_i^2\tau_i^2\boldsymbol{\Psi}_i^2\right), \quad \boldsymbol{\Psi}_i^2 = diag\left(\psi_{i,1}^2, ..., \psi_{i,k}^2\right), \tag{A.1}$$

$$\psi_{i,j} \sim Cauchy^{+}\left(0, 1\right), \quad j = 1, ..., k, \tag{A.2}$$

$$\tau_i \sim Cauchy^{+}\left(0, 1\right), \tag{A.3}$$

$$\boldsymbol{\Lambda}_{ij} \sim \begin{cases} N\left(0, \underline{h}_{ij}\right) I(\Lambda_{ij} > 0), & \text{if } S_{ij} = 1, \\ N\left(0, \underline{h}_{ij}\right) I(\Lambda_{ij} < 0), & \text{if } S_{ij} = -1, \\ \delta_0\left(\boldsymbol{\Lambda}_{ij}\right), & \text{if } S_{ij} = 0, \\ N\left(0, \underline{h}_{ij}\right), & \text{otherwise,} \end{cases} \quad j = 1, ..., r, \tag{A.4}$$

$$\mathbf{f}_t \sim N_r\left(\mathbf{0}, \mathbf{I}\right), \tag{A.5}$$

$$\sigma_i^2 \sim inv - Gamma\left(\underline{\rho}_i, \underline{\kappa}_i\right), \tag{A.6}$$

where we set $\underline{h}_{ij} = 4$, $\underline{\rho}_i = 1$ and $\underline{\kappa}_i = 0.01$.

Under these priors, the full factor sign restrictions algorithm takes the following form

Factor sign restrictions (FSR) algorithm

1. Sample $\boldsymbol{\phi}_i$ for $i = 1, ..., n$ from

$$\boldsymbol{\phi}_i|\boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \mathbf{f}, \mathbf{y} \sim N_k\left(\overline{\mathbf{V}}_i\left(\sum_{t=1}^{T}\sigma_i^{-2}\mathbf{x}_t'\widetilde{\mathbf{y}}_{it}\right), \overline{\mathbf{V}}_i\right), \tag{A.7}$$

   where $\widetilde{\mathbf{y}}_{it} = \mathbf{y}_{it} - \boldsymbol{\Lambda}_i\mathbf{f}_t$ and $\overline{\mathbf{V}}_i^{-1} = \left(\underline{\mathbf{V}}_i^{-1} + \sum_{t=1}^{T}\sigma_i^{-2}\mathbf{x}_t'\mathbf{x}_t\right)$. We use the efficient sampler of Bhattacharya et al. (2016) in order to sample these elements.

2. Sample ψ_{ij} using slice sampling (Neal, 2003)

   a. Set $\eta_{ij} = 1/\psi_{ij}^2$ using the last available sample of ψ_{ij}^2.

   b. Sample a random variable u from

$$u|\eta_{ij} \sim Uniform\left(0, \frac{1}{1+\eta_{ij}}\right). \tag{A.8}$$

c. Sample η_{ij} from

$$\eta_{ij} \sim e^{\frac{\phi_{ij}^2}{2\sigma_i^2}\eta_{ij}} I\left(\frac{u}{1-u} > \eta_{ij}\right) \tag{A.9}$$

and set $\psi_{ij} = 1/\sqrt{\eta_{ij}}$.

3. Sample τ_i using slice sampling (Neal, 2003)

a. Set $\xi_i = 1/\tau_i^2$ using the last available sample of τ_i^2.

b. Sample a random variable u from

$$v|\xi_{ij} \sim Uniform\left(0, \frac{1}{1+\xi_{ij}}\right). \tag{A.10}$$

c. Sample ξ_i from

$$\xi_i \sim \gamma\left((k+1)/2, v\frac{2\sigma^2}{\sum\left(\frac{\phi_{ij}}{\psi_{ij}}\right)^2}\right), \tag{A.11}$$

where $\gamma(\bullet)$ is the lower incomplete gamma function, and set $\tau_i = 1/\sqrt{\xi_i}$

4. Sample $\mathbf{\Lambda}_{ij}$ from univariate conditional posteriors (Geweke, 1996) of the form

$$\mathbf{\Lambda}_{ij}|\mathbf{\Lambda}_{-ij}\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{f}, \mathbf{y} \sim TN_{(\mathbf{a}_{ij}, \mathbf{b}_{ij})}\left(\overline{\lambda}_{ij} - \overline{h}_{ij}\sum_{l\neq j}\overline{h}_{il}^{-1}\left(\mathbf{\Lambda}_{il} - \overline{\lambda}_{il}\right), \overline{h}_{ij}\right), \tag{A.12}$$

where $\overline{\lambda}_{ij}$ and $\overline{h}_{ij}$ denote the ij^{th} elements of the joint posterior mean and variance, respectively, of $\mathbf{\Lambda}_i$. The joint posterior variance is $\overline{\mathbf{H}}^{-1} = \left(\underline{\mathbf{H}}^{-1} + \sum_{t=1}^{T}\sigma_i^{-2}\mathbf{f}_t'\mathbf{f}_t\right)$ and the joint posterior mean is $\overline{\mathbf{H}}\left(\sum_{t=1}^{T}\sigma_i^{-2}\mathbf{f}_t'\widehat{\mathbf{y}}_{it}\right)$ with $\widehat{\mathbf{y}}_{it} \equiv \boldsymbol{\varepsilon}_{it} = \mathbf{y}_{it} - \phi_i\mathbf{x}_t$. Here $TN_{(\mathbf{a}_{ij}, \mathbf{b}_{ij})}(\bullet)$ denotes the **univariate** truncated Normal distribution with bounds:

$$(\mathbf{a}_{ij}, \mathbf{b}_{ij}) = \begin{cases} (-\infty, 0) & \text{if } S_{ij} = -1, \\ (0, \infty) & \text{if } S_{ij} = 1, \\ (0, 0) & \text{if } S_{ij} = 0, \\ (-\infty, \infty) & \text{otherwise,} \end{cases} \tag{A.13}$$

We use the efficient univariate truncated Normal generator provided by Botev (2017) in order to sample these elements.

5. Sample $\mathbf{f}_t$ for $t = 1, ..., T$ from

$$\mathbf{f}_t|\mathbf{\Lambda}, \mathbf{\Sigma}, \mathbf{\Phi}, \mathbf{y} \sim N\left(\overline{\mathbf{G}}\left(\mathbf{\Lambda}\mathbf{\Sigma}^{-1}\widehat{\mathbf{y}}_t\right), \overline{\mathbf{G}}\right), \tag{A.14}$$

where $\overline{\mathbf{G}}^{-1} = (\mathbf{I}_r + \mathbf{\Lambda}'\mathbf{\Sigma}\mathbf{\Lambda})$. Post-process the draws of the $T \times r$ matrix $\mathbf{f} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$ such that its r columns (corresponding to structural shocks) are uncorrelated and standardized to unit variance. This is done by applying first the Gram-Schmidt procedure and subsequently dividing each column of $\mathbf{f}$ with its standard deviation.

6. Sample σ_i^2 for $i = 1, ..., n$ from

$$\sigma_i^2|\mathbf{\Lambda}, \mathbf{f}, \mathbf{\Phi}, \mathbf{y} \sim inv - Gamma\left(\frac{T}{2} + \underline{\rho}_i, \left[\underline{\kappa}_i^{-1} + \sum_{t=1}^{T}\left(\mathbf{y}_{it} - \phi_i\mathbf{x}_t - \mathbf{\Lambda}_i\mathbf{f}_t\right)'\left(\mathbf{y}_{it} - \phi_i\mathbf{x}_t - \mathbf{\Lambda}_i\mathbf{f}_t\right)\right]^{-1}\right) \tag{A.15}$$

B Data Appendix

B.1 Data for the simulation study

The data used in the simulation exercise are an augmented version of the data used in Arias et al. (2019). The measure of commodity prices they provide comes from Global Financial Data, while all remaining variables come originally from St Louis' Federal Reserve Economic Data (FRED; https://fred.stlouisfed.org/). Monthly GDP and GDP deflator are constructed using interpolation, and the reader is referred to the data supplement of Arias et al. (2019) for more information. The variables these authors use are augmented with stock prices, M1, unemployment rate, industrial production, employment, consumer prices (total), core consumer prices (total, less food and energy) and personal consumption deflator.

All variable mnemonics, short descriptions, and sources are visible in Table B1. The last column (Tcode) refers to the stationarity transformation used for each variable, namely 1: levels, and 5: first differences of logarithm (growth rates).

Table B1: Data used in Monte Carlo exercise

Mnemonic	Description	Source	Tcode
GDPC1	Monthly real GDP	Arias et al. (2019)	5
GDPDEFL	Monthly GDP deflator	Arias et al. (2019)	5
FEDFUNDS	Fed funds rate	Arias et al. (2019)	1
CPRINDEX	Commodity price index	Arias et al. (2019)	5
TRARR	Total reserves	Arias et al. (2019)	5
BOGNONBR	Nonborrowed reserves	Arias et al. (2019)	5
^GSPC	S&P 500 prices	Yahoo! Finance	5
M1REAL	Real M1 money stock	FRED	5
UNRATE	Unemployment rate	FRED	1
INDPRO	Industrial production index, all industries	FRED	5
PAYEMS	Employment, total	FRED	5
CPIAUSL	Consumer price index, all items	FRED	5
CPILFESL	Core CPI	FRED	5
PCEPILFE	Core PCE deflator	FRED	5

B.2 Data used for the large-scale VAR model for the US

The full list of variables is shown in Table B2 and they pertain to the sample 1985Q1 - 2013Q2. The original mortgage rate used by Furlanetto et al. (forthcoming) was only available after 1990Q1, and for that reason it has been replaced by the 30-year mortgage rate provided by FRED (contemporaneous correlation between the two series is 0.9967). All remaining series used in the empirical exercise are exactly those described in Table 11 of Furlanetto et al. (forthcoming), augmented with a few additional measures of output, consumer prices, and stock prices. The fifteen variables used in the large VAR can be seen in the first column of Table 4. In this list **stock prices** refers to S&P500 while **stock prices 2** refers to Dow Jones

Industrial Average. Similarly, **spread** refers to the Baa minus fed funds rate spread, while **spread 2** refers to the GZ credit spread.

Table B2: Data used in the empirical exercise

Variable	Description	Source
GDP	Log of real GNP/GDP	Federal Reserve Bank of Philadelphia
GDP deflator	Log of price index for GNP/GDP	Federal Reserve Bank of Philadelphia
Interest rate	3-month treasury bill	Federal Reserve Bank of St. Louis
Investment	Log of real gross private domestic investment	Federal Reserve Bank of St. Louis
Stock prices	Log of real S&P 500	Yahoo! Finance
Total credit	Log of loans to non-financial private sector	Board of Governors of the Federal Reserve System
Mortgages	Log of home mortgages of households and non-profit organizations	Board of Governors of the Federal Reserve System
Real estate value	Log of real estate at market value of households and non-profit organizations	Board of Governors of the Federal Reserve System
Corporate bond yield	Moody's baa corporate bond yield	Federal Reserve Bank of St. Louis
Federal funds rate	Federal funds rate	Federal Reserve Bank of St. Louis
GZ credit spread	Senior unsecured corporate bond spreads (non-financial firms)	Gilchrist and Zakrajšek (2012)[a]
EBP	Excess bond premium	Gilchrist and Zakrajšek (2012)[a]
VIX	Stock market volatility index	Bloom (2009)[b]
Mortgage rates	Home mortgages, fixed 30YR, Effective interest rate	Federal Reserve Bank of St. Louis
Employment	Log of total nonfarm employment	Federal Reserve Bank of Philadelphia
Core prices	Log of core consumer price index	Federal Reserve Bank of Philadelphia
Stock prices 2	Log of real DJIA	Yahoo! Finance

[a] Gilchrist, Simon, and Egon Zakrajšek (2012), Credit Spreads and Business Cycle Fluctuations. American Economic Review, 102 (4): 1692-1720.

[b] Bloom, Nicholas (2009), The Impact of Uncertainty Shocks. Econometrica, 77: 623-685.

C Additional results

C.1 Additional results for the first Monte Carlo exercise

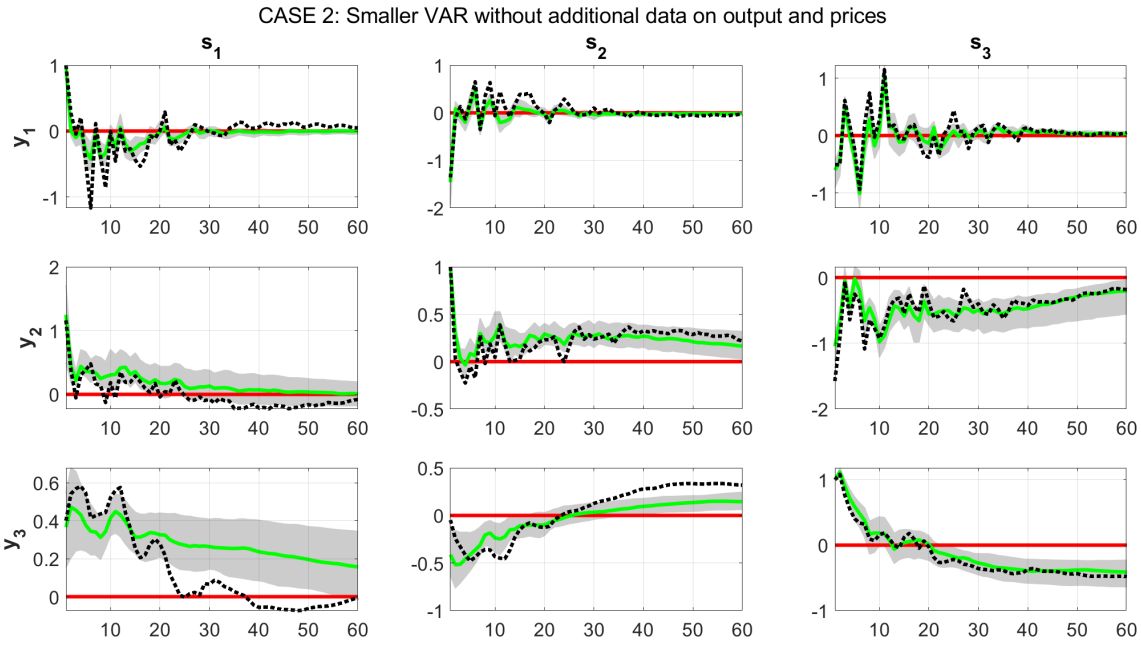


Figure C1: *Impulse response functions of the first three artificially generated variables (denoted as y_1, y_2, y_3) in response to the three identified shocks (denoted as s_1, s_2, s_3) in model C2 (misspecified VAR dimension). The green solid lines show the posterior median IRFs over the 500 Monte Carlo iterations, and the gray shaded areas their associated 90% bands. The true IRFs based on the DGP are shown using the black dashed lines.*

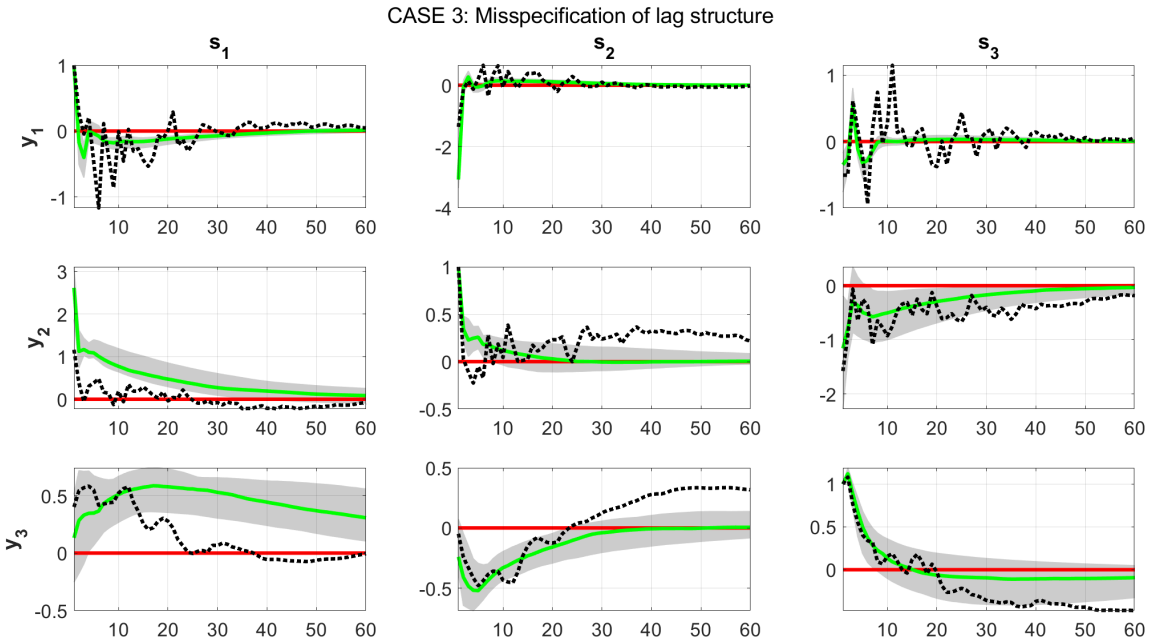


Figure C2: *Impulse response functions of the first three artificially generated variables (denoted as y_1, y_2, y_3) in response to the three identified shocks (denoted as s_1, s_2, s_3) in model C3 (misspecified number of lags). The green solid lines show the posterior median IRFs over the 500 Monte Carlo iterations, and the gray shaded areas their associated 90% bands. The true IRFs based on the DGP are shown using the black dashed lines.*

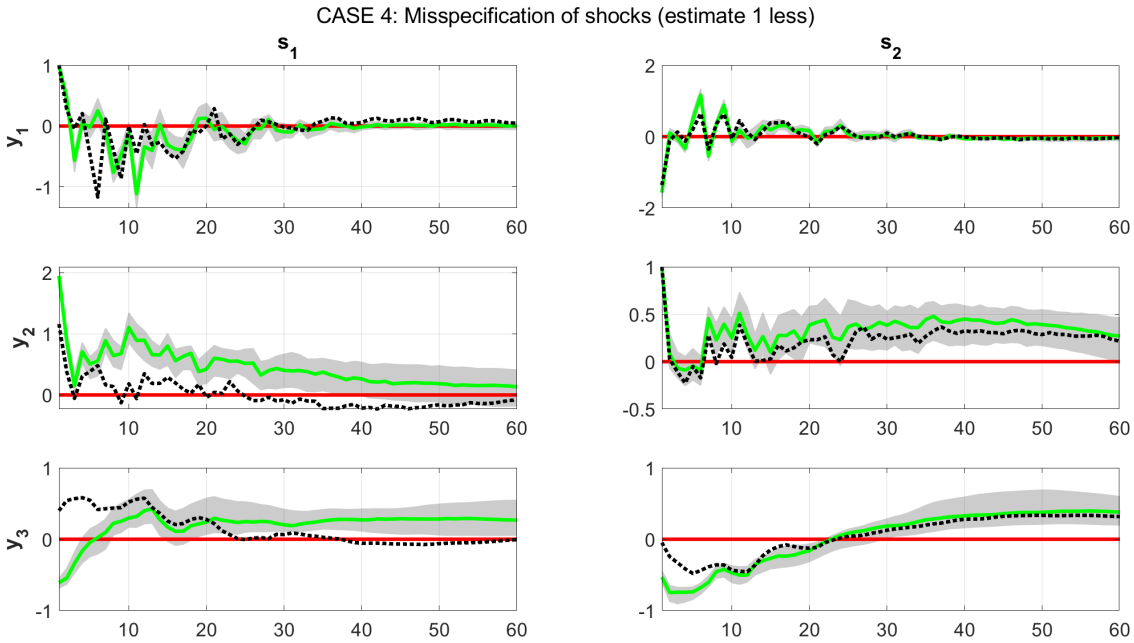


Figure C3: *Impulse response functions of the first three artificially generated variables (denoted as y_1, y_2, y_3) in response to the three identified shocks (denoted as s_1, s_2, s_3) in model C4 (misspecified number of shocks – one less). The green solid lines show the posterior median IRFs over the 500 Monte Carlo iterations, and the gray shaded areas their associated 90% bands. The true IRFs based on the DGP are shown using the black dashed lines.*

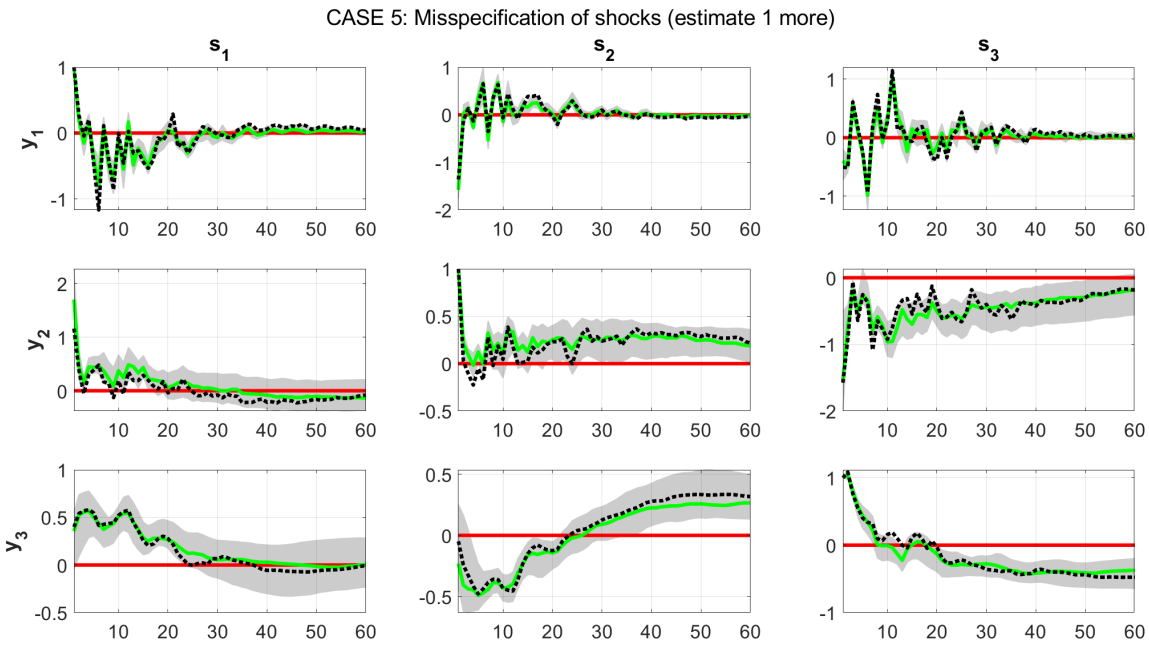
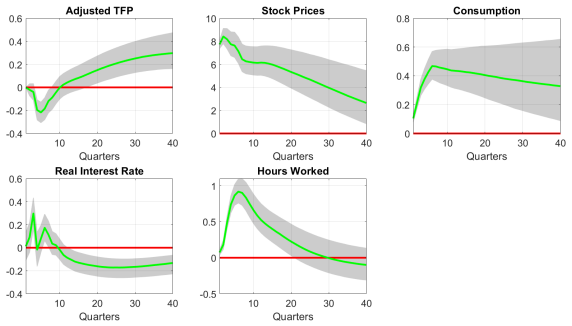


Figure C4: *Impulse response functions of the first three artificially generated variables (denoted as y_1, y_2, y_3) in response to the three identified shocks (denoted as s_1, s_2, s_3) in model C5 (misspecified number of shocks – one more). The green solid lines show the posterior median IRFs over the 500 Monte Carlo iterations, and the gray shaded areas their associated 90% bands. The true IRFs based on the DGP are shown using the black dashed lines.*

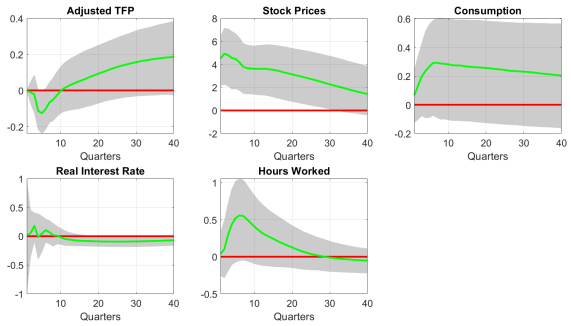
C.2 Additional empirical exercise: Measuring optimism shocks

Here I undertake an additional empirical exercise that will help shed more light on the performance of the new algorithm in real data. This empirical exercise is based on Section 6 of Arias et al. (2018). These authors use the example in Beaudry et al. (2011) in order to compare their novel importance sampling algorithm to the penalty function approach (PFA) of Mountford and Uhlig (2009). For the sake of comparability, we maintain their empirical setting, and for that reason we estimate VAR(4) models using the following five dependent variables: adjusted TFP, stock prices, consumption, the real interest rate, and hours worked. We identify a single optimism shock by restricting contemporaneously TFP to have a zero response, and stock prices to react positively. The signs in the remaining three variables are not restricted. Panels (a) and (b) in Figure C5 show the estimated responses using the PFA and importance sampling algorithms, respectively. These two panels are identical to panels

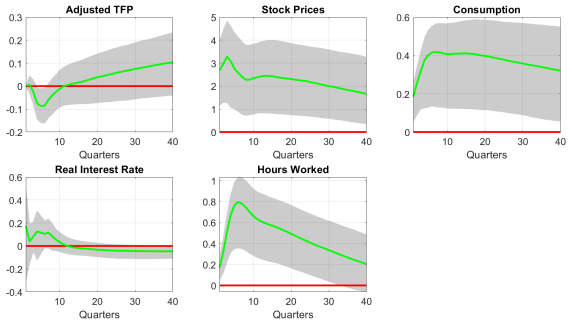
(a) and (b) in Figure 1 of Arias et al. (2018). The main point these authors make in their study is that the PFA algorithm ends up distorting the responses of stock prices, consumption and hours. Once their proposed importance sampling algorithm is considered, the significant responses found in Beaudry et al. (2011) disappear.



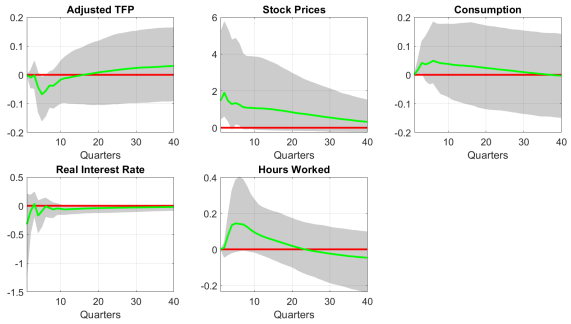
(a) Mountford and Uhlig (2009) algorithm



(b) Arias et al. (2018) algorithm



(c) Factor sign restrictions algorithm



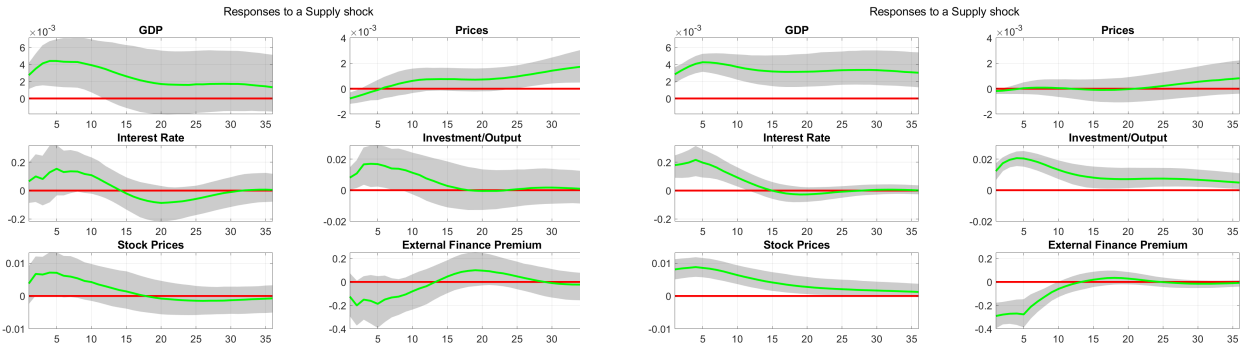
(d) Factor sign restrictions algorithm, with additional zero restrictions

Figure C5: *This figure replicates the results in Beaudry et al. (2011), using a five variable VAR for assessing the effects of an optimism shock. The sign-restricted impulse responses of the five variables are estimated using (a) the PFA algorithm of Mountford and Uhlig (2009); (b) the importance sampling approach of Arias et al. (2018); (c) the Gibbs sampler algorithm proposed in this paper; and (d) the algorithm of this paper, with additional zero restrictions in consumption and hours.*

Panel (c) of Figure C5 shows the results from the factor sign restrictions algorithm using the same zero restriction on TFP and positive sign restriction on stock prices. The response of stock prices is not as pronounced as in Beaudry et al. (2011), and in general the responses for TFP, stock prices and real interest rate are equivalent to Arias et al. (2018). However, the responses of consumption and hours are still strongly different from zero, even though

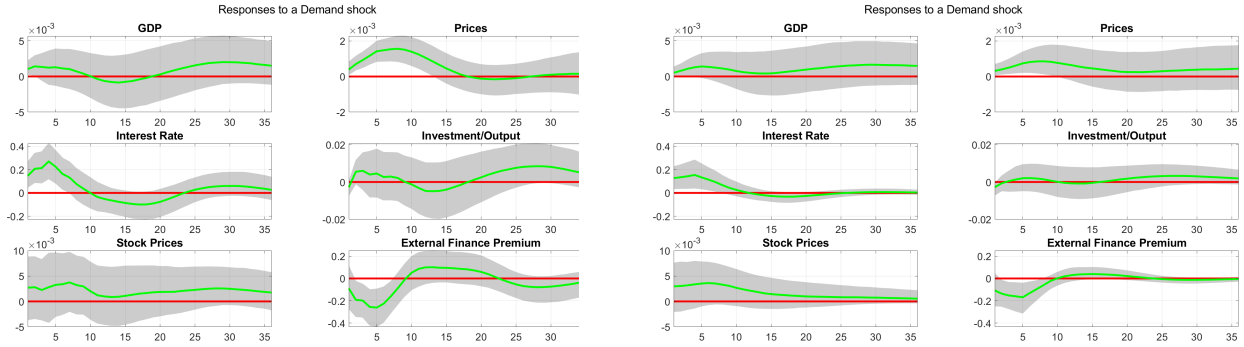
they have error bands and shapes that look much closer to those produced by the algorithm of Arias et al. (2018). Nevertheless, one aspect of the factor sign restrictions is that we can explicitly derive the implied fit to the VAR of imposing various restrictions. Therefore, we can explicitly test the premise of Arias et al. (2018) that consumption and hours are not affected by optimism shocks. Panel (d) in Figure C5 repeats estimation of the VAR using factor sign restrictions algorithm with additional zero restrictions in consumption and hours. The IRFs now look quantitatively and qualitatively closer to those in panel (b). Most importantly, we are able to test whether the model in panel (c) or (d) is supported by the data, that is, test whether the zero restrictions in consumption and hours. The DIC for the model without these restrictions is -13109.49 while the DIC for the model with the two zero restrictions is -15267.64. Thus, data evidence (which is conditional, of course, on the specific parametric likelihood specification and prior) suggests that the premise of Arias et al. (2018) – that optimism shocks do not affect consumption and hours – is correct.

C.3 Additional results for the baseline VAR of Furlanetto et al. (forthcoming)



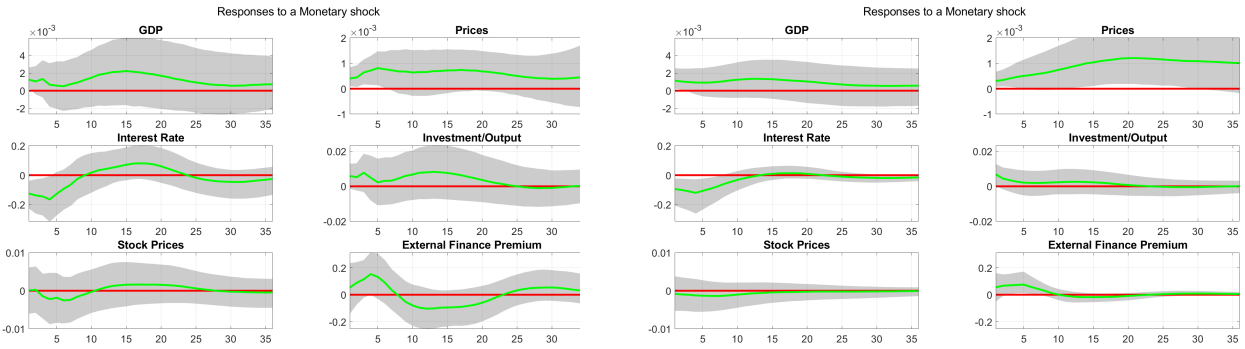
(a) Furlanetto et al. (forthcoming) algorithm (b) Factor sign restrictions algorithm

Figure C6: *This figure replicates the impulse response functions to an aggregate supply shock using the baseline specification of Furlanetto et al. (forthcoming). Panel (a) shows results based on the exact configuration of Furlanetto et al. (forthcoming, see Figure 1), using the algorithm of Rubio-Ramírez et al. (2010). Panel (b) replicates the same financial shock using the new sign restrictions algorithm.*



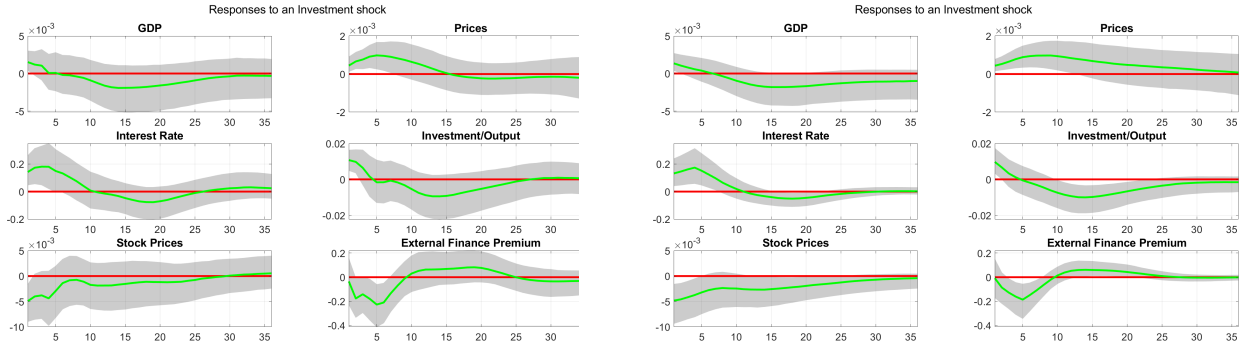
(a) Furlanetto et al. (forthcoming) algorithm (b) Factor sign restrictions algorithm

Figure C7: *This figure replicates the impulse response functions to an aggregate demand shock using the baseline specification of Furlanetto et al. (forthcoming). Panel (a) shows results based on the exact configuration of Furlanetto et al. (forthcoming, see Figure 1), using the algorithm of Rubio-Ramírez et al. (2010). Panel (b) replicates the same financial shock using the new sign restrictions algorithm.*



(a) Furlanetto et al. (forthcoming) algorithm (b) Factor sign restrictions algorithm

Figure C8: *This figure replicates the impulse response functions to a monetary policy shock using the baseline specification of Furlanetto et al. (forthcoming). Panel (a) shows results based on the exact configuration of Furlanetto et al. (forthcoming, see Figure 1), using the algorithm of Rubio-Ramírez et al. (2010). Panel (b) replicates the same financial shock using the new sign restrictions algorithm.*



(a) Furlanetto et al. (forthcoming) algorithm (b) Factor sign restrictions algorithm

Figure C9: *This figure replicates the impulse response functions to an investment shock using the baseline specification of Furlanetto et al. (forthcoming). Panel (a) shows results based on the exact configuration of Furlanetto et al. (forthcoming, see Figure 1), using the algorithm of Rubio-Ramírez et al. (2010). Panel (b) replicates the same financial shock using the new sign restrictions algorithm.*

C.4 Additional impulse responses from the large-scale, 15-variable VAR

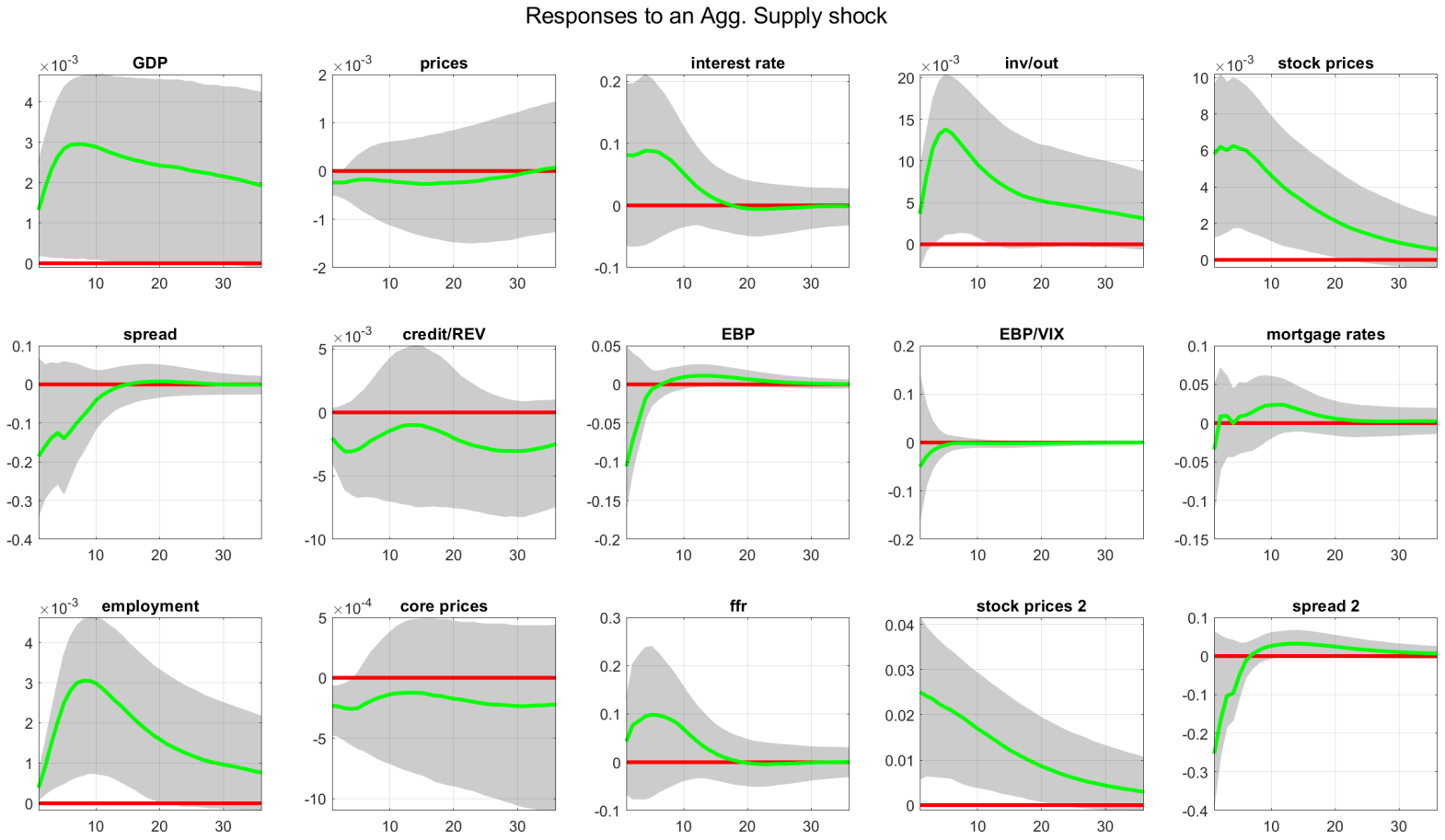


Figure C10: *Impulses response functions to an aggregate supply shock in the large, 15-variable VAR with seven shocks identified in total.*

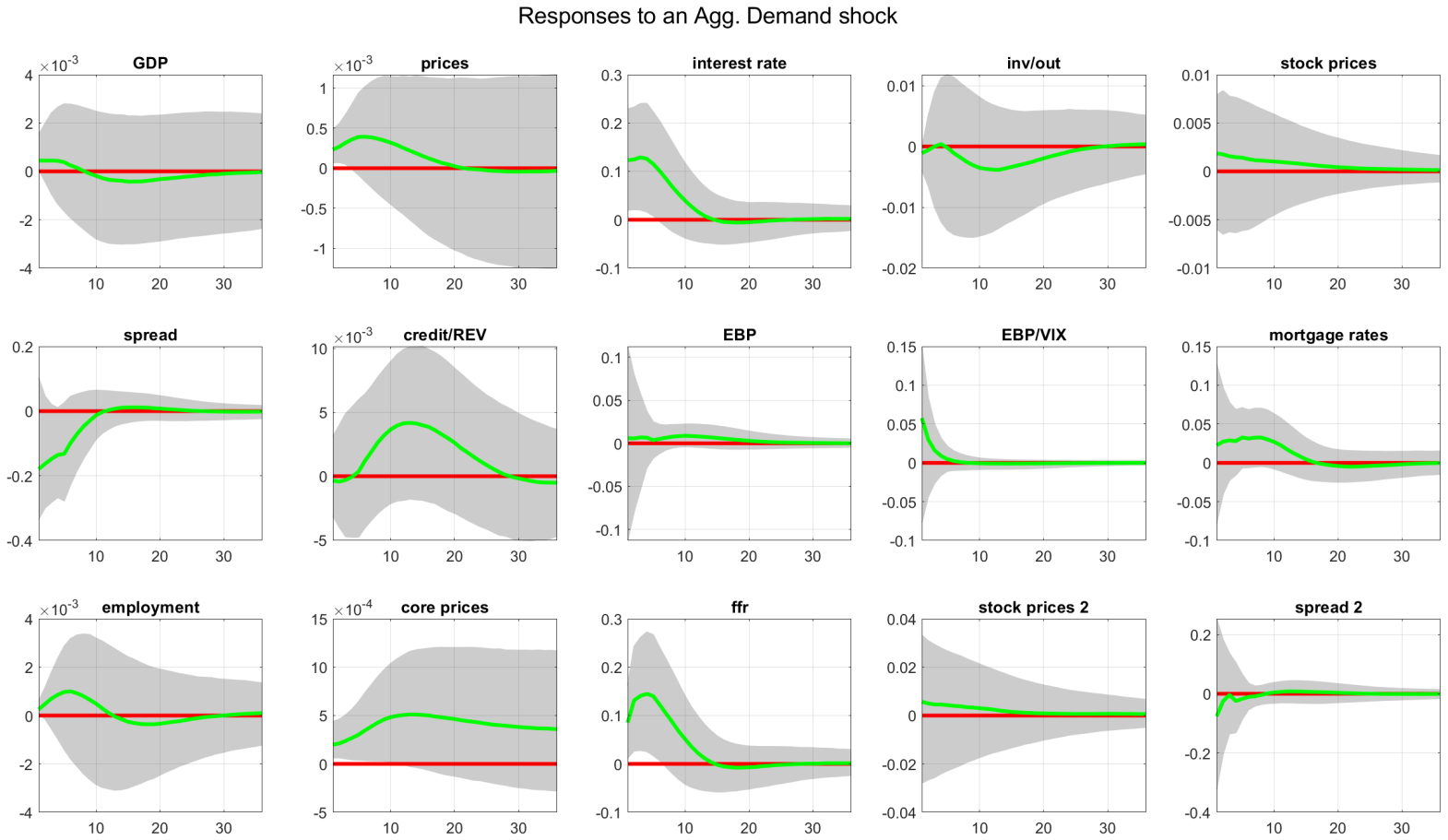


Figure C11: *Impulses response functions to an aggregate demand shock in the large, 15-variable VAR with seven shocks identified in total.*

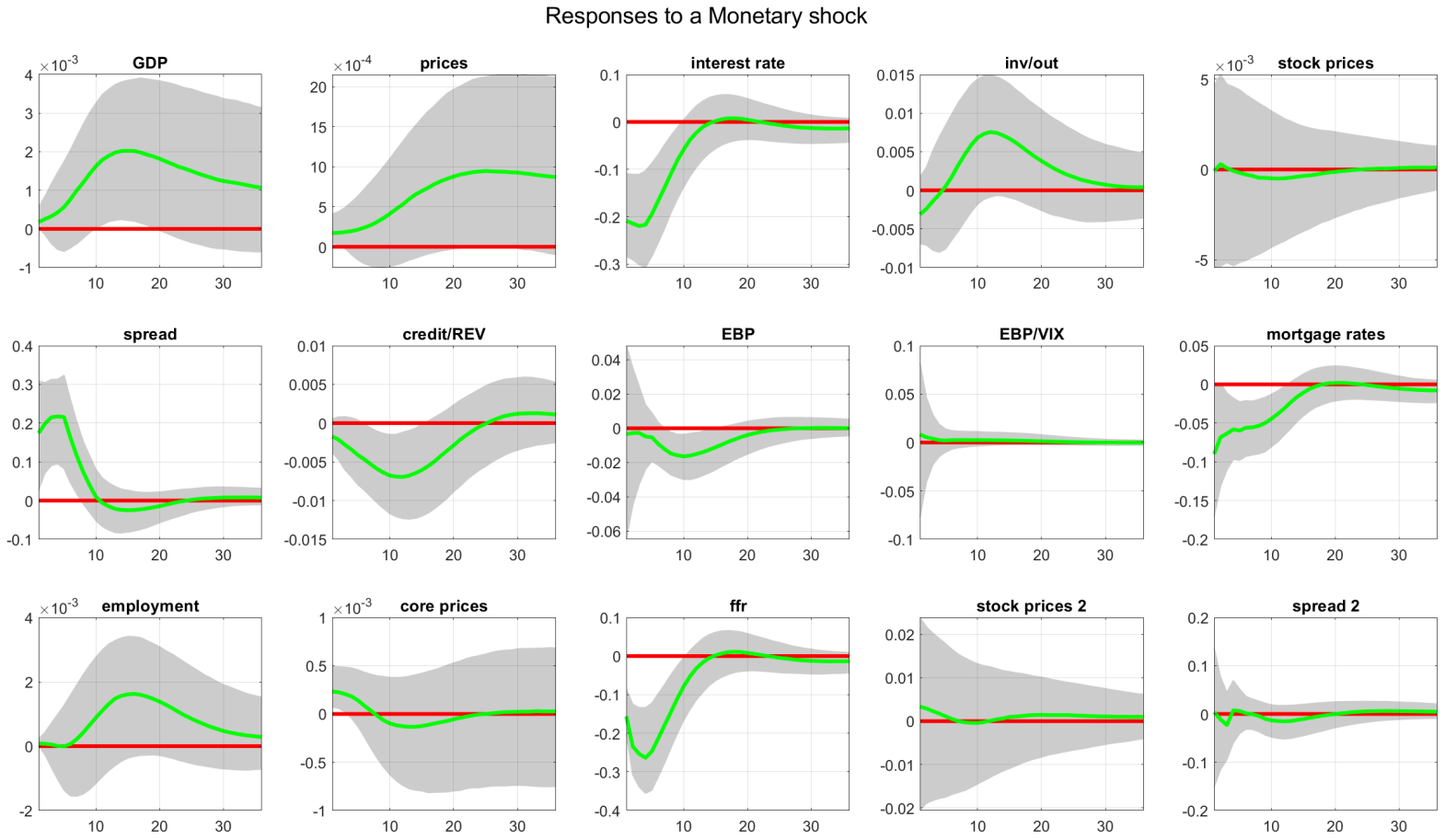


Figure C12: *Impulses response functions to a monetary policy shock in the large, 15-variable VAR with seven shocks identified in total.*

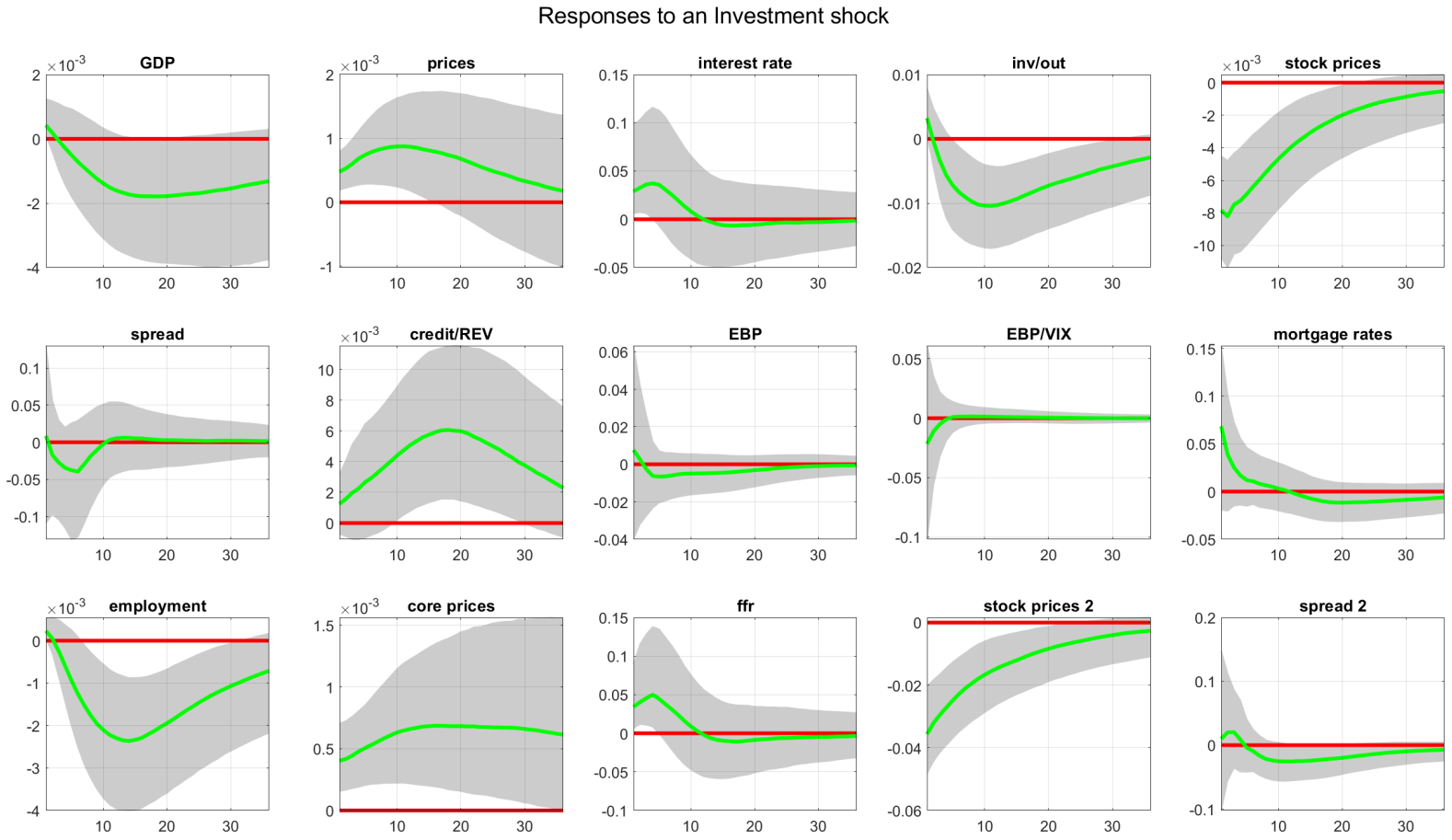


Figure C13: *Impulses response functions to an investment shock in the large, 15-variable VAR with seven shocks identified in total.*

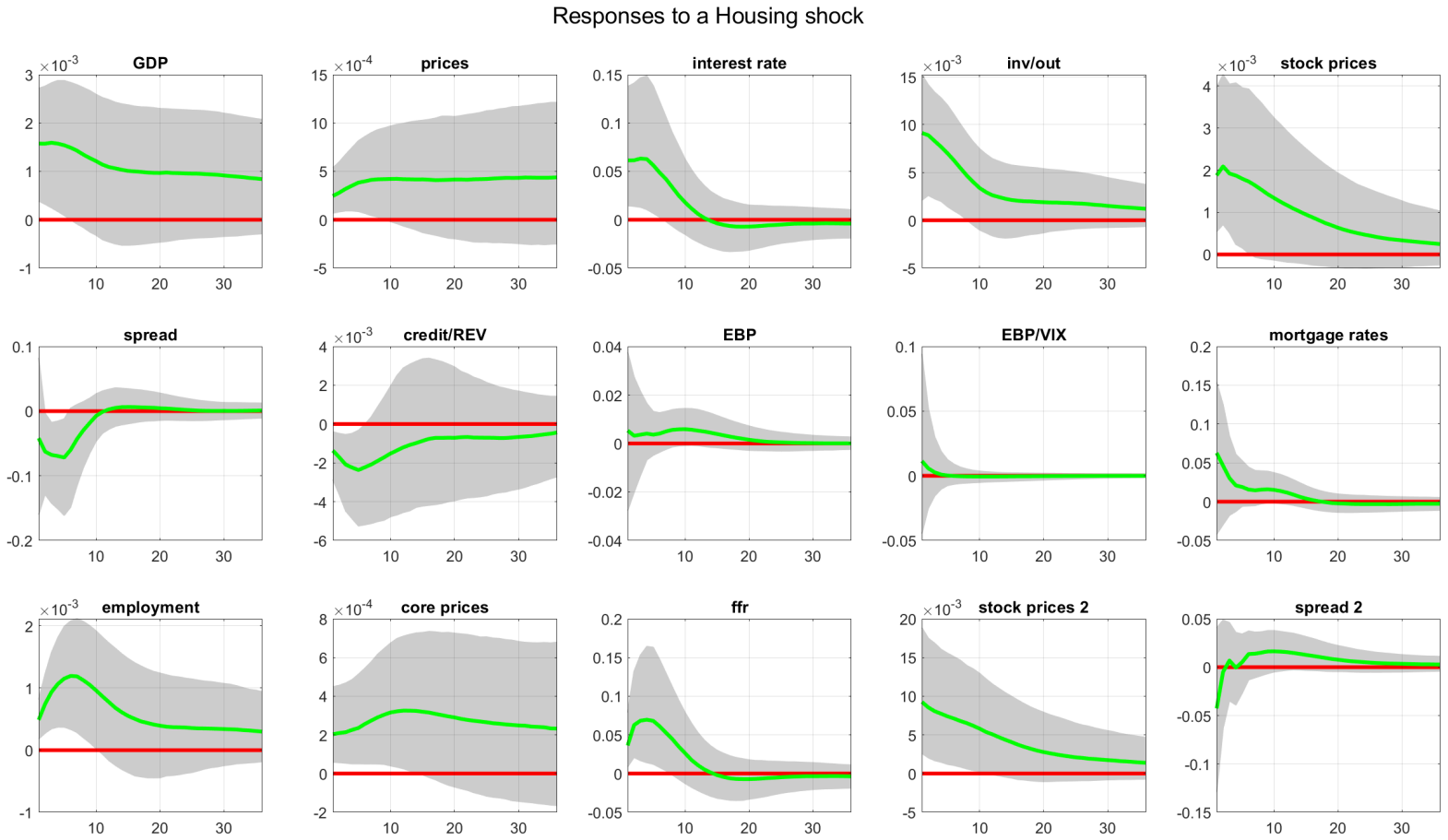


Figure C14: *Impulses response functions to a housing shock in the large, 15-variable VAR with seven shocks identified in total.*

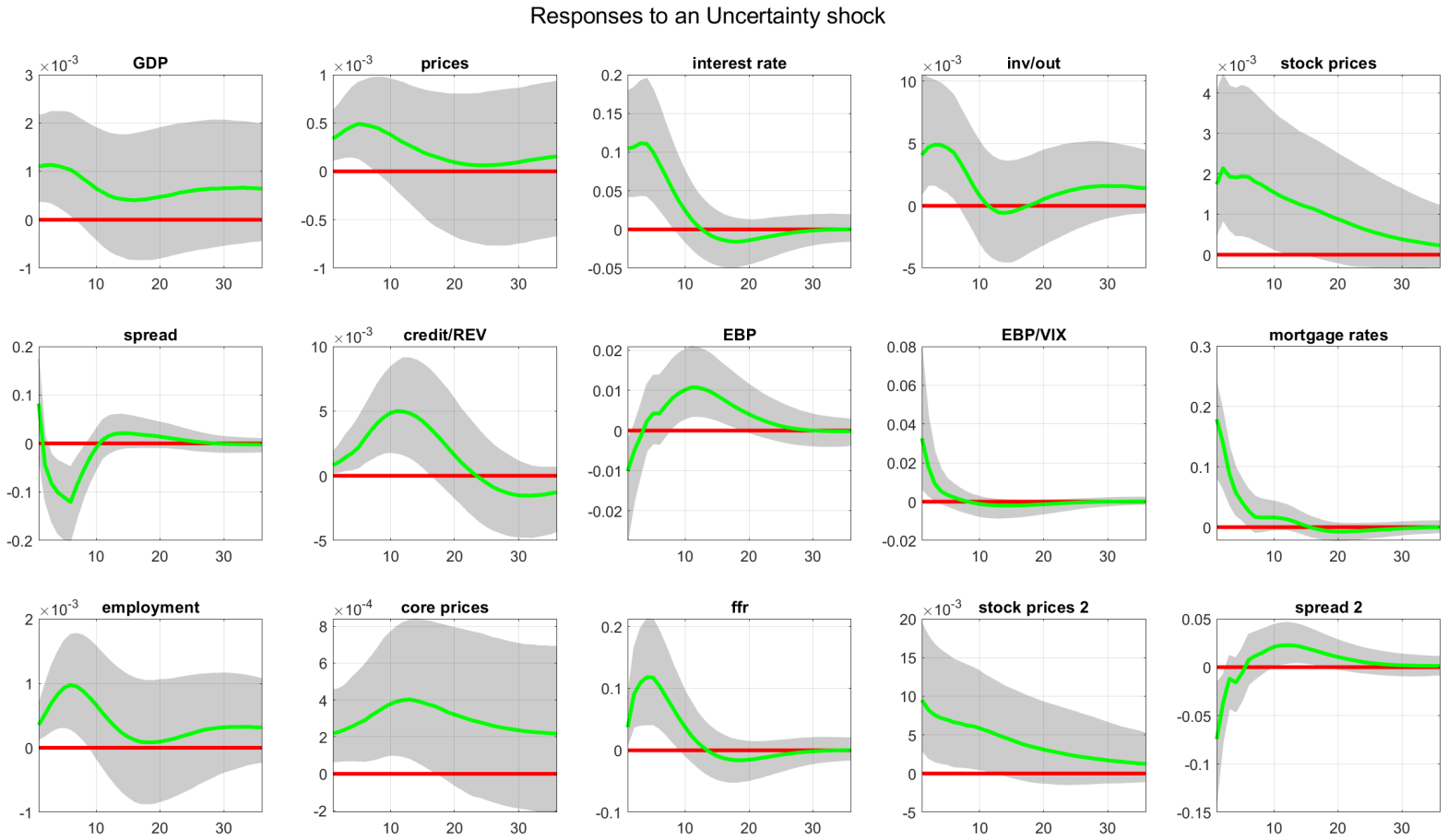


Figure C15: *Impulses response functions to an uncertainty shock in the large, 15-variable VAR with seven shocks identified in total.*