Identification Using Higher-Order Moments Restrictions

Philippe Andrade, Filippo Ferroni, and
Leonardo Melosi

August 18, 2023

WP 2023-28

https://doi.org/10.21033/wp-2023-28

FEDERAL RESERVE BANK of CHICAGO

Identification Using Higher-Order Moments Restrictions*

Philippe Andrade
FRB Boston

Filippo Ferroni
FRB Chicago

Leonardo Melosi
FRB Chicago, The University of Warwick & CEPR

August 18, 2023
First draft (June 2022)

Abstract

We exploit inequality restrictions on higher-order moments of the distribution of structural shocks to sharpen their identification. We show that these constraints can be treated as necessary conditions and used to shrink the set of admissible rotations. We illustrate the usefulness of this approach showing, by simulations, how it can dramatically improve the identification of monetary policy shocks when combined with widely used sign-restriction schemes. We then apply our methodology to two empirical questions: the effects of monetary policy shocks in the US and the effects of sovereign bonds spreads shocks in the Euro Area. In both cases, using higher-moment restrictions significantly sharpens identification. After a shock to EA governments bonds spreads, monetary policy quickly turns expansionary, corporate borrowing conditions worsen on impact, the real economy and the labor market of the Euro Area contract appreciably, returns on German government bonds fall, likely reflecting investors' flight to quality.

Keywords: Shock identification, skewness, kurtosis, VAR, sign restrictions, shocks to government bonds spreads, monetary shocks, Euro Area.

JEL classification: C32, E27, E32

*We would like to thank Tincho Almuzara, Fabio Canova, Raffaella Giacomini, Jonathan Wright, Lutz Kilian, Luca Gambetti and Giovanni Ricco and participants at various conferences and seminars for comments and inputs. We would like to thank Mary Tracy for her outstanding research assistance. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of Boston, the Federal Reserve Bank of Chicago or any other person associated with the Federal Reserve System.

1 Introduction

Since the seminal contribution of Sims (1980), numerous methods have been developed to study the propagation of unobserved unanticipated shocks to observed macroeconomic variables. Despite their differences, these identification schemes typically rely on second-order moment restrictions and on the assumption that the unconditional distribution of shocks is Gaussian.

However, macroeconomic outcomes often look non-Gaussian as normal times variations in aggregate variables coexist with unpredictable and unusually large positive or negative changes. Think of the recessions induced by the various types of financial crises (a stock market crash, a currency crisis, a sudden credit crunch, a sovereign debt crisis). Or the inflation surges induced by large supply disruptions. Or a regime shift in macroeconomic expectations induced by an abrupt and large change in monetary or fiscal policy. These rare unpredictable events can be viewed as being triggered by shocks much larger than that of ordinary shocks routinely perturbing the economy. In other words, macroeconomic outcomes can be seen as resulting from shocks that are drawn from a probability distribution with relatively fat tails of potentially asymmetric mass.

In this paper, we introduce a new method that exploits the information contained in the higher-order moments of the distribution of structural shocks to identify them. We postulate that the data generating process (DGP) of the economy is a vector autoregressive (VAR) model with non-gaussian shocks. This allows us to depart from gaussianity while keeping the simplicity of a linear transmission mechanism. We show how theoretically, restrictions on higher-order moments are necessary conditions to recover the true impact of non-gaussian structural shocks. We show how such restrictions can be implemented in practice, and how to deal with two issues that the departure from non-gaussianity raises, that is (i) the robustness of higher-order moments estimates and (ii) the identification and inference in VAR models with non-gaussian residuals. While our methodology can also be combined with other set-identifying conditions, such as sign, narrative or magnitude restrictions, we illustrate the usefulness of our approach by showing that imposing higher-order moment inequality restrictions on top of sign restrictions can improve considerably identification of monetary policy shocks in a medium-scale New-Keynesian model.[1] Finally, we apply the methodology to two empirical questions: the macroeconomic impact of US monetary policy shocks and the macroeconomic effects of sovereign spread shocks in the euro area.

We prove that these higher-order moment restrictions can be treated as necessary conditions; that is, it is impossible to obtain the "true impact" of the structural shock on the observable

[1] Moreover, we also show that in short samples the identification using higher-order moment inequality restrictions is more robust and reliable than point-identification schemes based on the spectral decomposition of the third and/or fourth moments.

variables without reproducing its correct higher-order moment. Therefore, all the rotations inconsistent with the economic *and* higher-moment restrictions can be mechanically discarded. The resulting set of rotations could be empty if, for instance, the higher moment inequality restrictions are inconsistent with the underlying data generating process. When it is not, the identified set is typically smaller as the rotations based on economic restrictions are always consistent with the second moments of the data but not necessarily with higher order moments. This occurs because orthonormal rotations while preserving covariances change the third and fourth moments of the data when different from the Gaussian implied ones. Furthermore, it is important to highlight that higher-order moment restrictions can also be imposed directly on the distribution of the observable variables' forecasts errors, provided that the structural shocks driving the true data generating process have zero coskewness and cokurtosis. For instance, this allows researchers to identify shocks that mostly contribute to the fat tails or the skewness of an observable variable of interest (e.g., risk premia or indicators of macro or policy uncertainty) at a specific horizon.

As sample estimates of third and fourth moments can be very sensitive to outliers and to minor perturbation of the data, we treat the distribution of the structural shock of interest non-parametrically and impose restrictions on the distance between different percentiles of the empirical distribution to generate the desired asymmetry and/or fat-tailedness.[2] We propose to use a Bayesian approach to estimate the VAR reduced-form parameters with non-Gaussian errors. The Bayesian approach builds on the work by Petrova (2022), where she proposes a robust and computationally fast Bayesian procedure which delivers asymptotically correct posterior credible sets without the need for distributional assumptions.

We use our identification approach to study the dynamic propagation of policy and non policy macroeconomic shocks. In a first application, motivated by the fact that the observed measures of monetary policy surprises feature significant departures from Gaussianity (see section 4.1), we use excess kurtosis restrictions to study the transmission of (conventional) monetary policy in the U.S. to output and inflation. In particular, to identify monetary policy shocks we postulate that interest rate and prices move in opposite directions and leave unrestricted the impact on real activity; we assume further that these shocks are leptokurtic. With these assumptions, we consistently find that a monetary policy tightening induces a significant contraction in output.

In a second application, we use our identification strategy to a VAR model estimated using macro-financial data of the Euro Area (EA), including the spread between Italian and German government bonds, to study the transmissions of sovereign spread shocks. In particular, we

[2]Loosely speaking, robust estimators of the kurtosis for example consider the ratio between the distance of the percentiles in the tails and the distance between the percentiles close to the median. In the case of leptokurtic shocks, the larger the numerator, the thicker the tails; the smaller the denominator the more clustered is the distribution around the median.

postulate that the spread shock is positively skewed with fat-tails, increases on impact the 10 year Italian government bond yield and widens the Italian German five year government bond yield differential. An increase in spread of one percent generates a flight to quality to Germany with a decline in the ten year yield of about 40 bps. Credit conditions tighten immediately and borrowing costs increase more than one-to-one to the increase in the sovereign spread. The real economy of the EA also responds on impact after the spread shock and these effects are re-absorbed in less than three years after the shock. The effects on the unemployment rate of the EA are quite more severe and longer lasting.

This application allows us to illustrate that higher-moment restrictions considerably sharpens our ability to isolate the macroeconomic effects of large and abrupt spread shocks relative to traditional sign restrictions. We find that rapid increase in spreads have serious implications for the real economy of the EA with a sharp increase in corporate borrowing costs and an abrupt increase in the unemployment rate. EA monetary policy eases up considerably to combat the recessionary consequences of these non-Gaussian spread shocks. The fall in the unemployment rate and the response of monetary policy are considerably less pronounced when one entirely relies on traditional identification schemes, such as Cholesky orthogonalization or sign restrictions, which do not impose any restrictions on higher-order moments of the shocks. This finding makes sense because a "normal" change to the spread like the ones we observe every day arguably does not trigger a massive change in the monetary policy stance, nor is it expected to have major consequences for the cost of borrowing, the real economy, and the labor market.

The paper is organized as follows. Next section (1.1) discusses the existing literature. Section 2 presents our identification approach. Section 3 shows the role of kurtosis in identifying the effect of monetary policy on output in standard NK models. Section 4 and 5 illustrate the usefulness of our approach in two empirical applications: measuring the real effects of the US conventional monetary policy and the impact of sovereign spread shocks in the euro area. Section 6 concludes.

1.1 Literature review

Our paper relates to the literature that exploits the properties of non-Gaussian distributions of empirical innovations to obtain statistical identification of shocks; in particular, point identification is achieved by assuming some functional non-Gaussian distributions for the VAR reduced form errors, see Lanne, Meitz and Saikkonen (2017) and Gouriéroux, Monfort and Renne (2017, 2019), or Jarociński (2021). Differently from this literature, our methodology does not make any distribution assumptions about the sources of randomness and do not assume a specific distribution of the structural shock of interest. Other scholars achieve statistical identification of the structural shocks by minimizing the distance between the VAR model-implied empirical innovation higher-order moments and sample counterparts, e.g. Lanne, Liu and Luoto (2022) among others. As higher-order moments of the empirical innovations are often difficult to estimate in short samples (see section 3.2 for an example when the DGP is a standard medium-scale New-Keynesian model), our methodology does not rest on sample estimates of the empirical innovation higher-order moments. We impose inequality restrictions on the higher-order moments of the structural shock itself using non-parametric methods based on the distance between different percentiles of the shock's empirical distribution; this approach turns out to be more robust in short samples. Importantly, differently from previous literature, this generates a set-identification of the shock on interest (as opposed to the system point identification). This in turn allows to couple these restrictions with other popular macroeconomic assumptions based on the signs (or zero) impact of the shock on the endogenous variables, magnitude or elasticity bounds or narrative restrictions on historical episodes. In this respect our approach is modular.

Our method can also be seen as a way to sharpen the identification of structural shocks using sign-restrictions. Indeed, in our applications, we combine sign restrictions with higher-order moment restrictions. Kilian and Murphy (2012), Arias, Rubio-Ramírez and Waggoner (2018), Wolf (2020, 2022) show that imposing sign restrictions alone is often too weak to provide adequate identification of structural shocks. In this paper, we show that exploiting higher-order information can sharpen the identification of non-Gaussian shocks achieved by imposing sign restrictions. Other refinements of the sign-restriction approach has been proposed recently in the literature (e.g. Antolin-Diaz and Rubio-Ramírez 2018). Most of these resolutions rest on bring in external instruments or narrative evidence to sharpen the identification based on sign restrictions. Instead the proposed methodology only exploits information that is internal to our model. This external information may not be available very far back in the past, shortening the sample period considerably.

Our paper is related to the work by Drautzburg and Wright (2023) where they propose a robust frequentist approach to narrow the identification set of structural shocks. They refine the identification set by discarding rotations that are not consistent with statistical independence of the structural shocks and look at higher-order moments of candidate shocks to enforce that. However, the use of the higher-order moment restrictions is different from what we do. Specifically, their methodology rules out shocks whose higher moments are not consistent with independence. We do not impose that. In contrast, we require that a specific structural shock has important non-Gaussian features. Moreover, their approach requires computing the higher order moments (or the marginal empirical distribution) of all the structural shocks in the system which grows with the dimension of the VAR. Our restrictions are constructed on a handful of higher moments and do not depend on the number of endogenous variables in the VAR.

2 Identification with higher moments

To better isolate the identification problem we abstract momentarily from estimation issues and assume that the econometrician observes a vector of *empirical innovations*, ι_t, where ι_t is a $n \times 1$ vector of *uncorrelated* innovations with unit variance, i.e. $\iota_t \sim (0, I)$. These empirical innovations can be thought as the orthogonalized reduced form residuals of a VAR model obtained from a long time series of data. Importantly, we assume that the empirical innovations are combinations of n unobserved structural shocks, ν_t, with unit variance and non-trivial third or fourth moments, i.e. $E(\nu_{i,t}^3) \neq 0$ or $E(\nu_{i,t}^4) \neq 3$ for some i. Moreover, we assume that $E(\nu_{i,t}\nu_{j,t}\nu_{k,t}) = 0$ for all $i \neq j, k$, $E(\nu_{i,t}\nu_{j,t}\nu_{k,t}\nu_{n,t}) = 0$ for all $i \neq j, k, m$ and $E(\nu_{i,t}^2\nu_{j,t}^2) = 1$ for all $i \neq j$. Finally, there is a linear mapping between empirical innovations and structural shocks, i.e.

$$\iota_{1,t} = \alpha_{1,1}\nu_{1,t} + ... + \alpha_{1,m}\nu_{n,t}$$
$$\vdots$$
$$\iota_{n,t} = \alpha_{n,1}\nu_{1,t} + ... + \alpha_{n,n}\nu_{n,t}$$

and we define with A_o the matrix collecting the structural coefficients $\alpha_{i,j}$. Since both ι_t and ν_t are uncorrelated with unit variance, it must be that $A_o'A_o = A_oA_o' = I$ and the following equations hold $\iota_t = A_o\nu_t$ and $A_o'\iota_t = \nu_t$. When ν_t is a Gaussian distributed random vector, moments higher than the second are not useful for identification since third moments are zero and fourth moments are invariant to orthonormal rotations, see appendix A.2.

Typically, applied researchers identify one or a subset of structural shocks with economic restrictions, i.e. by imposing signs, magnitude or zero restrictions on the elements of the orthonormal rotation matrix A.[3] However, the identified set of admissible A's remains in many situations very large; see section 3 for an example. In what follows, we show that this set can be narrowed further if the structural shocks of interest have non-trivial higher moments and if these restrictions are used for identification. In other words, while the rotation matrix A does not change second moments, it does change the moments higher than the second.

The higher order moments have been used in Independent Component Analysis (ICA) for reconstructing the original (demixed) sources of variations from a vector of mixed signals. The core result of this literature is that if at most one of the components ν_t is Gaussian, A is point identified up to sign change and permutation of its columns; therefore, all the structural shocks can be reconstructed from the empirical innovations. This result is shown by Comon (1994, Theorem 11) and has been discussed and used for identification in the literature, see also

[3]In the VAR and dynamic factor model contexts, restrictions are also imposed on the endogenous variable at particular horizons (long run as well) and not necessarily only on impact. In this section we abstract from this. However, it does not have any consequence on the conclusions derived in this section.

Gouriéroux et al. (2019) and Lanne et al. (2017).

Regardless of the number of non-guassian shocks, statistical point-identification of the impact of the structural shock of interest can be derived using the full array of higher order moments of the empirical innovations. Notice that we can express the $(n \times n^2)$ matrix collecting the third moments as follows

$$E(\nu_t \nu_t' \otimes \nu_t') = \sum_{i=1}^{n} \zeta_i J_i \otimes \boldsymbol{e}_i'$$

where ζ_i represents the third moment of the structural shocks i, $\boldsymbol{e}_i$ is the $n \times 1$ vector with zeros everywhere except a one in the i^{th} position, J_i the $n \times n$ matrix of zeros everywhere except one in the i^{th} position of the main diagonal. It is easy to show that the squared third moment matrix is a $(n \times n)$ diagonal matrix

$$E(\nu_t \nu_t' \otimes \nu_t') E(\nu_t \nu_t' \otimes \nu_t')' = \left(\sum_{i=1}^{n} \xi_i J_i \otimes \boldsymbol{e}_i \right) \left(\sum_{i=1}^{n} \xi_i J_i \otimes \boldsymbol{e}_i \right)' = \Lambda_\zeta$$

where Λ_ζ is a diagonal matrix collecting the squared third moments of the structural shocks.[4] The matrix collecting the empirical innovation third moments can be expressed as $E(\iota_t \iota_t' \otimes \iota_t') = A_o E(\nu_t \nu_t' \otimes \nu_t')(A_o \otimes A_o)'$ and the matrix collecting the squared third moments of the empirical innovations is given by

$$E(\iota_t \iota_t' \otimes \iota_t') E(\iota_t \iota_t' \otimes \iota_t')' = A_o \Lambda_\zeta A_o'$$

which gives rise to a eigenvalue/eigenvector or spectral decomposition. In particular, the eigenvalue corresponds to the square of the third moments of the structural shock and the corresponding unit-length eigenvector coincides with the column of the original mixing or impact matrix, up to a sign switch and permutation of columns. As long as the shock of interest has non-zero third moment, we can identify its impact on the empirical innovations using the full array of empirical innovations third moments.

Similar arguments carry over for forth moments. In fact, we can express the $(n^2 \times n^2)$ matrix collecting the fourth moments *in excess of* the standard normal ones as follows

$$E(\nu_t \nu_t' \otimes \nu_t' \otimes \nu_t) - \mathcal{K}^z = \sum_{i=1}^{n} \xi_i J_i \otimes J_i$$

where ξ_i is the excess kurtosis of the structural shock i and $\mathcal{K}_n$ is the matrix of fourth moments of a normal standard multivariate distribution. The latter is a diagonal matrix with non-zero elements only on the positions $j(n+1) - n$ for $j = 1, ..., n$. The matrix collecting the fourth moments of the empirical innovations, ι_t, *in excess of* the standard normal ones can be expressed

[4] See appendix A.2 for more details.

as

$$E(\iota_t\iota_t' \otimes \iota_t' \otimes \iota_t) - \mathcal{K}^z = (A_o \otimes A_o)(E(\nu_t\nu_t' \otimes \nu_t' \otimes \nu_t) - \mathcal{K}^z)(A_o \otimes A_o)'$$
$$= P\Lambda_\xi P'$$

where Λ_ξ is a diagonal matrix where the first largest n eigenvalues corresponds to the excess kurtosis of the structural shocks and the matrix P collects the corresponding eigenvectors. One can derive the column of the original rotation matrix by taking the first n elements of the first n eigenvectors divided by the absolute value of the first elements of the eigenvector, i.e. $P(1:n,j)/\sqrt{|P(1,j)|}$ for $j = 1, ..., n$.[5] Up to a permutation of columns and a sign switch, the resulting matrix coincides with the original impact matrix.

It is important to highlight two points. First, structural shocks need to have zero cross third and fourth moment. As in structural models disturbances are typically assumed to be statistically independent, the latter represents a reasonable assumption. Second, even if one is interested in identifying the impact of one single shock, one needs to compute the full set of the third or fourth moments of the empirical innovations (or their consistent estimates) to retrieve the column of interest of the rotation matrix. And the number of moments grows significantly with the number of the empirical innovations, i.e. with the size of the VAR model. Moreover and more importantly, estimates of the fourth or third sample moments can be very sensitive to outliers or minor perturbation of the data and their estimates might be imprecise in short samples. This in turns might lead to corrupted estimates of the shock's impact on observables; we showe this point in section 3.2 using as laboratory a standard macro New-Keynesian model.

Rather than achieving structural identification from estimates of the the empirical innovations higher moments, our preferred approach is to impose inequality restrictions on the higher moments of the structural shock itself. Our approach requires only estimating one single moment (i.e. the skewness or kurtosis) of the structural shock of interest for which we can easily compute robust estimators based on the percentile of the empirical distribution of the shock, e.g. see appendix A.4; thus avoiding computing a large number of sensitive sample third or fourth moments. Moreover, the inequality restriction imposes weaker conditions generating set-identification as opposed to point-identification and it can be coupled with other assumptions, such as signs, zeros, narrative, magnitude and/or statistical independence restrictions. In this sense higher-moments inequality restrictions are *modular*, as they can be flexibly used standalone or in combination with other assumptions.

There exists an important relationship between the higher moments of the structural shock and the column that measures its impact on the empirical innovations; this relationship allows us to discard matrices that do not generate the desired properties of the shock of interest. This

[5] Alternatively, one can consider the procedure outlined in Kollo (2008) which is described in the appendix.

result is shown more formally in the following propositions. Without loss of generality assume that we are interested in the last (n^{th}) shock, and let $\boldsymbol{\alpha}_n$ bet the rightmost (n^{th}) columns of A_o (the true impact matrix), which measure the impact of the structural shocks on the empirical innovations.

Proposition 1 *Let $\breve{\nu}_{n,t} = \boldsymbol{a}_n' \iota_t$ be a candidate structural shock with $\boldsymbol{a}_n$ a unit-length vector of weights. When $\boldsymbol{a}_n = \boldsymbol{\alpha}_n$ where $\boldsymbol{\alpha}_n$ is the 'true' impact of the structural shock on the empirical innovations, then candidate shock ($\breve{\nu}_{n,t}$) has the same higher-order property of the 'true' structural shock.*

The proof is in the appendix A.2. The proposition offers a necessary condition for identification. Consider a rotation matrix that violates the higher moment condition, that is $E(\breve{\nu}_{n,t}^4) \neq \xi_n$ or $E(\breve{\nu}_{n,t}^3) \neq \zeta_n$, then according to the propositions it must be the case that $\boldsymbol{a}_n \neq \boldsymbol{\alpha}_n$. In other words it is impossible to get the true rotation matrix without generating the correct higher moments. This implies that in the estimation procedure we can discard all the rotations that are inconsistent with the postulated higher order moments of the structural shock because these rotations are inconsistent with the true impact of the structural shock on the empirical innovations. Clearly, since we do not directly observe structural shock we typically ignore their higher order properties. However, economic insight and measured proxies might inform us on the nature of the higher moments of a particular macro shock. For example, in section 4.1 we show that the typical observed measures of monetary policy shocks are leptokurtic. While we might ignore the exact amount of 'taildeness', we have estimates that points at values statistically and significantly larger than the normal distribution and use those as reference values. Similarly, one could argue that uncertainty measured by the market-implied volatility could be driven by an underlying process whose distribution has a thick left tail so that extreme negative events are more likely. Therefore, positive skewness might be assumed.

When higher moments restrictions have sufficient grounds for justification, then these restrictions might be useful for identification. Since the rotation matrix modifies the higher moments of the candidate identified shock, the set of admissible rotations can be reduced by imposing the desired higher moments to lie in a preassigned interval. Clearly, if the 'true' moments of the structural shocks of interest are inconsistent with the restriction, the set of admissible rotations is empty. E.g., if we assume that monetary policy shocks are leptokurtic while they are not in the true data generating process, we will not find rotations satisfying the restrictions. It is important to highlight that similar considerations apply also to sign, magnitude or narrative restrictions; if we impose 'incorrect' restriction on –say– the IRF of the structural shocks either the set of accepted rotation is empty or the identification of the shock is corrupted because it does not reflect the sign pattern of the data generating process. On the contrary, this class of restrictions allows to shrink the set of rotations when the underlying data generating process is consistent with the postulated restrictions.

Estimates of fourth and third sample moments (i.e. kurtosis and skewness) can be very sensitive to outliers with short samples, see Kim and White (2004). In this paper we propose to use a robust non-parametric approach to compute asymmetries and tailedness. These robust estimators are constructed using the empirical distribution of the structural shocks of interest. For fourth moments or tailedness, robust estimators considers the ratio between the distance of the percentiles in the tails and the distance between the percentiles close to the median. For example, the larger the numerator, the thicker the tails; the smaller the denominator the more clustered is the distribution around the median and hence the more leptokurtic the shock is. For third moments or asymmetries, robust measures exploit the standardized distance between median and mean. In the appendix (section A.4) we discuss some of these robust estimators more in details. In what follow we define asymmetries and taildeness with

$$\mathcal{S}(x) = \frac{F^{-1}(0.5) - \bar{x}}{\text{std}(x)}, \quad \mathcal{K}(x) = \frac{F^{-1}(0.975) - F^{-1}(0.025)}{F^{-1}(0.75) - F^{-1}(0.25)} \tag{1}$$

where $F^{-1}(\alpha)$ is the α-percentile of the empirical distribution of x.

More formally, our identification and estimation approach can be described as follows. We assume that the observed data are generated by a $VAR(p)$ model,

$$y_t = \Phi_1 y_{t-1} + ... + \Phi_p y_{t-p} + \Phi_0 + u_t$$

where y_t is $n \times 1$ vector of endogenous variables, Φ_0 is a vector of constant and Φ_j are $n \times n$ matrices. We assume $y_0, \ldots, y_{-p+1}$ are fixed. We assume that u_t are i.i.d. zero mean random vectors with unconditional covariance matrix Σ. We assume that the VAR reduced form shocks are linear combination of the unobserved structural shocks, ν_t, i.e.

$$u_t = \Sigma^{1/2} \iota_t = \Sigma^{1/2} \ \Omega \ \nu_t$$

where $\Sigma^{1/2}$ is the Cholesky factorization of Σ and Ω is an orthonormal matrix, i.e. $\Omega\Omega' = \Omega'\Omega = I$. The structural shocks, ν_t, are zero-mean orthogonal shocks with unitary variance, i.e. $\nu_t \sim (0, I)$. Standard inference on VAR parameters typically postulates a multivariate normal distribution for the reduced form innovations. Such an assumption cannot be considered in our context. We propose to adopt a robust Bayesian approach which allows to construct posterior credible sets without the need for distributional assumptions of the reduced form residuals. The Bayesian approach we use builds on the work by Petrova (2022), where she propose a robust and computationally fast Bayesian procedure to estimate the reduced form parameters of the VAR in the presence of non-Gaussianity. The robust approach relies on the asymptotic normality of the Quasi Maximum Likelihood (QML) estimator[6] of reduced form parameters,

[6]The QML is the maximum estimator of the quasi-likelihood. The quasi-likelihood in this context coincides with the likelihood of the VAR when *incorrectly* assuming normality of the reduced form residuals.

autoregressive coefficients and covariances; Petrova (2022) derives the closed form expression for the asymptotic covariance matrix of the QML estimator allowing for fast simulation from its asymptotic distribution. Combined with a prior, one can draw the VAR reduced form parameters from the posterior distribution without specifying the shock distribution.[7] Assume that we are interested in identifying the last shock, $\nu_{n,t}$, let $\Sigma^{(j)}$ and $\Phi^{(j)}$ be the j^{th} draw from the reduced form parameters poterior, we can identify the structural shock using higher-order restrictions as follows.

- Draw $\breve{\Omega}$ from a uniform distribution with the Rubio-Ramírez, Waggoner and Zha (2010) algorithm and

  I. compute the impulse response function and check if the sign (or any other economic) restrictions are verified

  II. compute the implied structural shocks

$$\breve{\nu}_t^{(j)} = \breve{\Omega}' \left(\Sigma^{(j)}\right)^{-1/2} (y_t - \Phi_1^{(j)} y_{t-1} - \dots - \Phi_p^{(j)} y_{t-p} - \Phi_0^{(j)})$$

  II. compute $\mathcal{S}(\breve{\nu}_{n,t}^{(j)})$ and $\mathcal{K}(\breve{\nu}_{n,t}^{(j)})$ and check if the higher-order moment inequality restrictions are satisfied

  If both [I] and [III] are satisfied, keep the draw $\Omega^{(j)} = \breve{\Omega}$. Else repeat [I], [II] and [III].

After a suitable number of iterations, the draws are representative of the posterior distribution of the impulse responses of interest. The estimation of the reduced form parameters and the computation of the impulse responses using the higher-order moments is performed using the toolbox described in Ferroni and Canova (2021).[8]

2.1 Analytical example

In this section we offer an analytical example to illustrate the main point of the paper, that is inequality restrictions on higher moments of the structural shock can lead to restrict the identified set of the structural parameter. Assume that $n = 2$ and that $\nu_{1,t} \sim (0,1)$ and $\nu_{2,t} \sim (0,1)$. Moreover, we assume that the third moments of the structural shocks are zero and one respectively, i.e. $E(\nu_{1,t}^3) = 0$ and $E(\nu_{2,t}^3) = 1$, and and cross third moments are zero, i.e. $E(\nu_{1,t}\nu_{2,t}) = E(\nu_{1,t}\nu_{2,t}^2) = E(\nu_{1,t}^2\nu_{2,t}) = 0$. The structural equations are

$$\iota_{1,t} = \cos\theta_o \nu_{1,t} - \sin\theta_o \nu_{2,t}$$

$$\iota_{2,t} = \sin\theta_o \nu_{1,t} + \cos\theta_o \nu_{2,t}$$

[7] See appendix A.5 for more details on the estimation.

[8] Codes for replication can be found on the Github page.

where θ_o is the 'true' unknown angle of rotation with $\theta_o \in (-\pi/2, \pi/2)$ and $A_o A_o' = A_o' A_o = I$. The econometrician does not observe ν_t and only observes ι_t. It is trivial to notice that first and second moments do not depend on θ_o; hence it is not possible to retrieve A_o. However, third moments do. In particular, the population third moments of the observed empirical innovations are given by $E(\iota_{1,t}^3) = -\sin^3\theta_o$, $E(\iota_{2,t}^3) = \cos^3\theta_o$, $E(\iota_{1,t}^2\iota_{2,t}) = \sin^2\theta_o\cos\theta_o$, and $E(\iota_{1,t}\iota_{2,t}^2) = -\sin\theta_o\cos^2\theta_o$. Using the insights of the previous section, we can retrieve the impact matrix from the third moments of the empirical innovations. First notice that

$$\begin{aligned} E(\iota_t\iota_t' \otimes \iota_t') &= E\left(\begin{pmatrix} \iota_{1,t}^2 & \iota_{1,t}\iota_{2,t} \\ \iota_{1,t}\iota_{2,t} & \iota_{2,t}^2 \end{pmatrix} \otimes \begin{pmatrix} \iota_{1,t} & \iota_{2,t} \end{pmatrix}\right) \\ &= \underbrace{\begin{pmatrix} \sin^2\theta_o & -\sin^2\theta_o\cos\theta_o \\ -\sin\theta_o\cos\theta_o & \cos^2\theta_o \end{pmatrix}}_{\Omega} \otimes \begin{pmatrix} -\sin\theta_o & \cos\theta_o \end{pmatrix} \end{aligned}$$

Since Ω is an idempotent matrix, i.e. $\Omega\Omega = \Omega$, we have that

$$E(\iota_t\iota_t' \otimes \iota_t')E(\iota_t\iota_t' \otimes \iota_t')' = \Omega$$

The characteristic polynomial of Ω is $(\sin^2\theta_o - \lambda)(\cos^2\theta_o - \lambda) - \sin^2\theta_o\cos^2\theta_o$ and the associated eigenvalues are zero and one respectively. This means that the first structural shock third moment equals zero and the second structural shock third moment equals one. The eigenvector associated with the non-zero eigenvalue is $\begin{pmatrix} -\sin\theta_o & \cos\theta_o \end{pmatrix}'$.

Our preferred approach does not use the third moments of the empirical innovations directly and imposes weaker restrictions. In particular, we assume that the second shock has positive skewness. Moreover, we do not make assumptions about the third moment of the other shock. Yet, we can restrict considerably the identified set. More formally, we only consider the set of rotations such that the following inequality is verified,

$$E(\breve{\nu}_{2,t}^3) > 0$$

Let A be a generic rotation with angle θ, i.e. $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ and let $\breve{\nu}_t = A'\iota_t$. The corresponding population moment is then

$$\begin{aligned} E(\breve{\nu}_{2,t}^3) &= E(-\sin\theta\iota_{1,t} + \cos\theta\iota_{2,t})^3 = [\sin\theta\sin\theta_o + \cos\theta\cos\theta_o]^3 \\ &= [\text{sign}(\cos\theta_o)\cos(\theta - \theta_o)]^3 > 0 \end{aligned}$$

Since $\text{sign}(\cos\theta_o) = 1$ for all $\theta_o \in (-\pi/2, \pi/2)$, the latter is positive whenever $\cos(\theta - \theta_o) > 0$, which occurs when in the region of points where[9]

$$\max\{-\pi/2, \theta_o - \pi/2\} < \theta < \min\{\pi/2, \theta_o + \pi/2\}$$

[9]More details on the solution can be found in the appendix A.3.1.

This condition allows to shrink the set of admissible rotations which otherwise would be the set $\theta \in (-\pi/2, \pi/2)$; in particular, when θ_o is positive (negative), the lower (upper) bound shrinks. There is a discontinuity at $\theta_o = 0$. In such a case, the first (second) structural shock coincides with the first (second) empirical innovation, all third moments are zero but the third moment of the second empirical innovation, $E\iota^3_{2,t}$, and there is no point in considering them as a system of mixed signals.

Figure 1 reports graphically the regions just described, i.e. the combination of values of θ where the third moment of $\breve{\nu}^3_{2,t}$ is positive for any given value of θ_o. The white areas indicate the region where the third moment is non-positive. The gray-colored area denotes positive values for the third moment. The black solid line indicates the region where $\theta = \theta_o$ and where the third moment equals one. As we move far way from the points where $\theta = \theta_o$ the third moment of shock of interest decreases, up to the point where it turns negative. This occurs in the north-west and south-east corners.

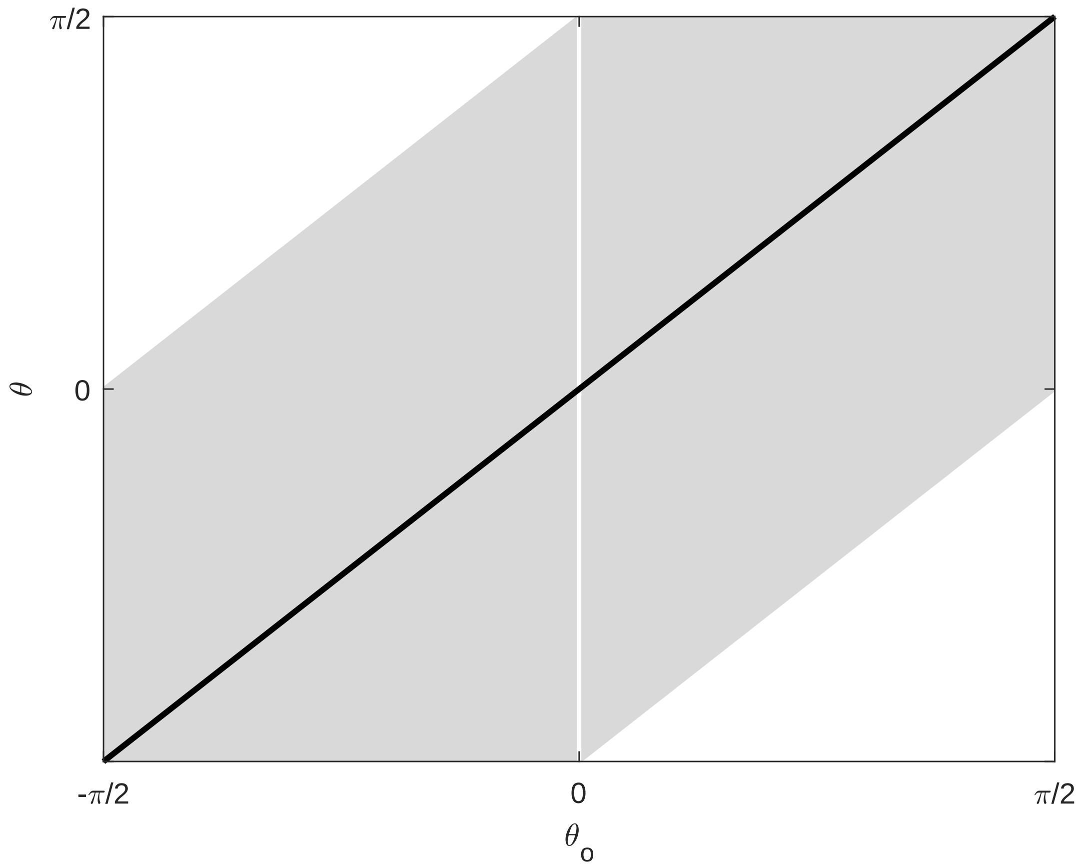


Figure 1: Values of $E(\breve{\nu}^3_{2,t})$ for different combinations of θ (y-axis) and θ_o (x-axis). Gray region indicates non negative values.

3 The effects of monetary policy shocks on output

Inequality restrictions on higher moments are not only important from a statistical viewpoint as they sharpen our identification sets as we discussed in the previous section. They are also economic meaningful as they can, in certain situations, better isolate the effect of economic shocks in our standard macro models.

With an agnostic identification procedure Uhlig (2005) finds that monetary policy shocks have no clear effect on output. He imposes sign restrictions on inflation and interest rate (moving in opposite directions) and is *agnostic* about the response on output. Wolf (2020) shows that Uhlig (2005)'s result is consistent with the standard New Keynesian (NK) model and this occurs because supply and demand shocks tend to *masquerade* or disguise as monetary policy shocks when only sign restrictions on inflation and interest rate are imposed. He concludes that pure sign restrictions are quite weak identifying information. Identification can be improved with instruments or restriction on the reaction coefficients of the policy function as in Arias, Caldara and Rubio-Ramírez (2019).

Higher moments inequality restrictions can also resolve the *masquerading* shocks problem when the monetary policy shock has fourth moment sufficiently different from supply and demand. We show this result using a three equations NK model and a more realistic model with a variety of shocks and frictions in the same spirit as Christiano, Eichenbaum and Evans (2005) or Smets and Wouters (2007).

3.1 NK model

We consider the simplest static version of the NK model. Detailed derivations of the conventional three-equation New Keynesian model are offered in Galí (2015). The model in its log linearized form is described by three equations,

$$
\begin{aligned}
y_t =& y_{t+1|t} - (i_t - \pi_{t+1|t}) + \sigma_d \epsilon_t^d \\
\pi_t =& \beta \pi_{t+1|t} + \kappa y_t - \sigma_s \epsilon_t^s \\
i_t =& \phi_\pi \pi_t + \phi_y y_t + \sigma_m \epsilon_t^m
\end{aligned}
$$

y is real output; i is the nominal interest rate (the federal funds rate); and π is inflation. The model has three structural disturbances: a demand shock ϵ_t^d, a supply shock ϵ_t^s, and a monetary policy shock ϵ_t^m. The first equation is a standard IS-relation (demand block), the second equation is the New Keynesian Phillips curve (supply block), and the third equation is the monetary policy rule (policy block). It is straightforward to show that this benchmark model is static and admits the closed-form solution:

$$x_t = \begin{pmatrix} y_t \\ \pi_t \\ i_t \end{pmatrix} = \frac{1}{1+\kappa\phi_\pi+\phi_y} \begin{pmatrix} \sigma_d & \phi_\pi\sigma_s & -\sigma_m \\ \kappa\sigma_d & -(1+\phi_y)\sigma_s & -\kappa\sigma_m \\ (\phi_y+\kappa\phi_\pi)\sigma_d & -\phi_\pi\sigma_s & \sigma_m \end{pmatrix} \begin{pmatrix} \epsilon_t^d \\ \epsilon_t^s \\ \epsilon_t^m \end{pmatrix} = A_o\epsilon_t$$

Unlike standard practice, we depart from the assumptions of normality of shocks and postulate that the monetary policy has positive excess kurtosis. We consider Gaussian supply and demand shocks and leptokurtic monetary policy shocks. In particular, we assume that

$$\epsilon_t^d \sim N(0,1),\ \epsilon_t^s \sim N(0,1),\ \epsilon_t^m \sim Laplace(0,1).$$

The excess kurtosis of the Laplace (Gaussian) distribution is 3 (0). We assign the following values to the parameters $\sigma_s = \sigma_d = \sigma_m = 1$, $\phi_\pi = 1.5$, $\phi_y = 0.5$ and $\kappa = 0.2$ and simulate a long time series of data, $T = 100,000$, compute the sample covariance of the data and generate candidate rotations using the Haar prior (see Rubio-Ramírez et al. (2010)).

The masquerading effect discussed in Wolf (2020) is displayed in left panel of figure 2 which reports the scatter plot of all the supply and demand realizations with blue circles and with red circles the scatter plot of the combination of realizations that generate a negative comovement of inflation and interest rate. The latter occurs when supply shocks are relatively small, i.e. the support of demand $(-4, 4)$ is roughly double the support of supply $(-2, 2)$. Moreover, the scatter plot also reveals an upward sloping relationship between supply and demand, meaning that the realizations tend to have the same sign. If we focus on positive realizations, while supply pushes inflation down, the endogenous part of the monetary policy rule responds to demand pressures prescribing an interest rate hike; hence, this combination of demand and supply generates a pattern similar to a policy tightening. However, unlike a monetary policy tightening, the large demand expansion and the mildly positive supply drive output up.

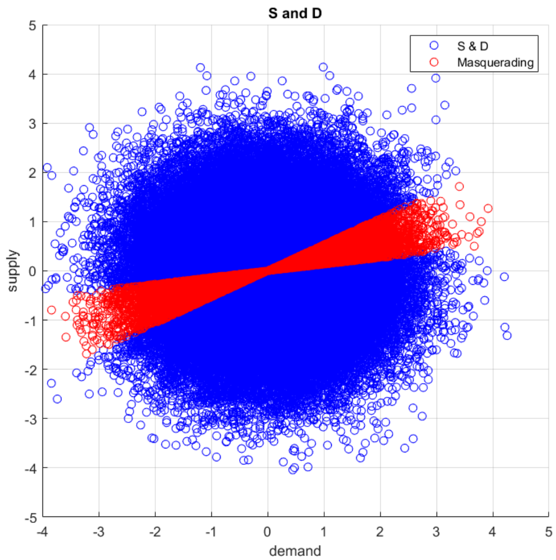
As a result, if one imposes sign restrictions only on inflation and interest rate, the impact on output can be both positive and negative because the masquerading effect of supply and demand shocks. This is visible in the right panel of figure 2 where the blue bars report the distribution of the impact on output of monetary policy shocks identified with sign restriction only. We notice immediately that the support can be positive and negative.

When we complement the sign restrictions with a moment restriction on the excess kurtosis of monetary policy shock, the impact distribution changes. In particular, we postulate that the excess kurtosis of monetary policy shocks needs to be larger than a threshold value k, i.e. $E(\check{\epsilon}_t^4) - 3 > k$, and we consider different values for the this threshold. When we impose that the excess kurtosis of monetary policy shocks ought to be larger than 1.5, almost all the positive responses of output are chopped away and the distribution tilts towards negative outcomes only as indicated by the red-colored bars in the right top panel of figure 2.

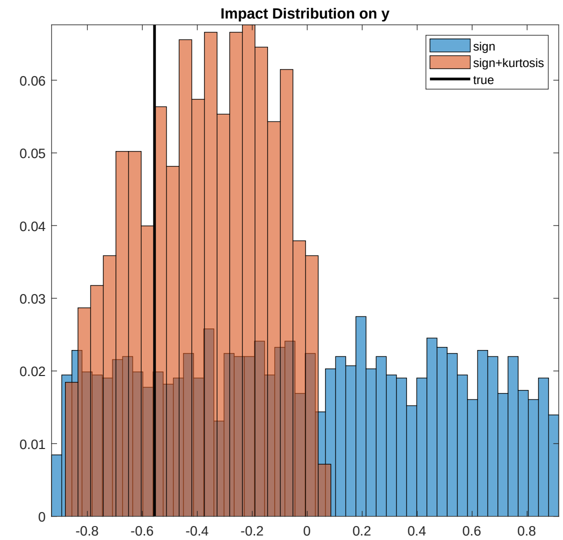
To see why this occurs, it is useful to think of the following experiment. When the standard deviation of the monetary policy shock becomes very small, $\sigma_m \rightarrow 0$, the fluctuations in the data are driven only by supply and demand shocks, which are both Gaussian distributed. Any linear combination of Gaussian distributed random variables is Gaussian random variable; therefore, its excess-kurtosis will not be different from zero. In the bottom panel of figure 2 we report distribution of the excess-kurtosis of the masquerading shock across different rotations, where we can see that the excess-kurtosis is not different from zero. The restriction on monetary policy shock being leptokurtic mechanically discards all these candidate rotations, which generate a negative correlation between inflation and interest rate but a positive correlation between output and inflation. Therefore, the sign and excess-kurtosis restrictions together can single out the correct response of output after an interest rate hike as shown in right top panel of figure 2.

Clearly the bound for the excess kurtosis are crucial to shrink the identified set. A lower threshold for the kurtosis of the monetary policy shock would increase the support of the impact distribution on y, while a higher threshold would work in the opposite direction of shrinking the distribution around the true negative value. The role of this threshold is reported in figure 2 panel (d) where we plot with a blue line the probability of positive response of output against different values of k, the lower bound of the excess kurtosis interval. The figure reports also the acceptance rate when only considering sign restrictions (grey line) and the acceptance rate when using signs and excess kurtosis restrictions (red line). When $k = 0$, roughly one fourth of rotations are accepted with sign only and with sign and kurtosis restrictions, respectively; the probability that the response of output is positive is roughly one half. As we increase the minimum value for the excess kurtosis, fewer rotations verify both the signs and the excess kurtosis restrictions, and the probability of a positive response of output to a monetary policy shock declines. Eventually it becomes zero when we impose that the excess kurtosis of monetary policy shocks needs to be larger than 1.8.

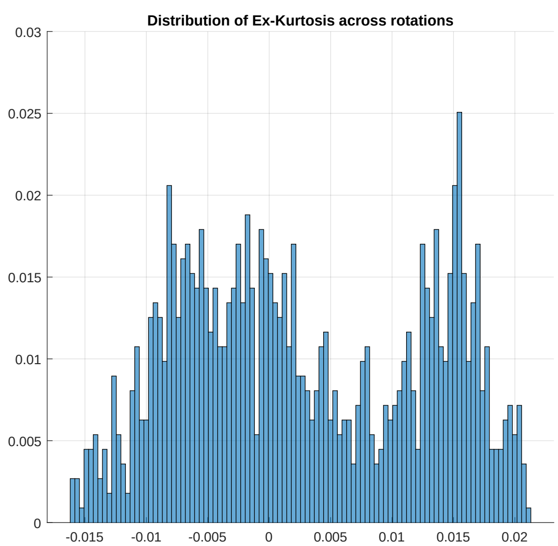
In the next section, we explore the validity of these conclusions in a more realistic model which features a number of real and nominal frictions.



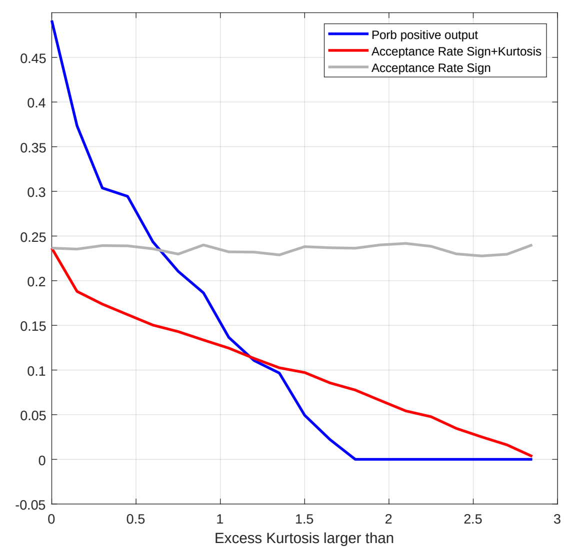
(a) Realizations of demand and supply shocks: all (blue circles) and masqueraded MP (red circles).



(b) Distribution of MP impact on output with sign and with moment and sign restrictions.



(c) Distribution of the excess-kurtois of the masqueraded shock across different rotations.



(d) Probability of positive response of y at different restrictions on monetary policy excess kurtosis.

Figure 2: Masquerading effect (a); impact distribution of monetary policy on output (b); distribution of excess-kurtosis for the masquerading shock; and (d) probability of positive response of y after a monetary policy tightening.

3.2 Smets and Wouters (2007) model

The Smets and Wouters (2007) (SW) model is perhaps the most well-known example of an empirically successful structural business cycle model. In this model the fluctuations in economic activity, labor market variables, prices and interest rate are explained by handful of shocks; these are technology, risk premium, investment demand, monetary policy, government/exogenous spending, and price and wage markups shocks. We use this model as a laboratory to show how the higher order moments can sharpen identification. In most of the analysis here we consider the SW posterior mode parameterization and the smoothed estimates of the shocks using postwar US data on output, consumption, investment, real wages, inflation, interest rate and hours worked as in their original work.[10]

As in the simple NK model version, also in the SW model demand and supply shocks can generate in combination a sign pattern for inflation and interest rate similar to that produced by a monetary policy shock; hence the transmission of monetary policy shocks might be difficult to isolate using only sign restrictions. Figure 3 reports –from top to bottom– the impulse response of technology (supply), risk premium (demand) and monetary policy shocks to output, inflation and interest rate in the SW model respectively; the last raw displays the dynamic transmission of the sum of supply and demand where the responses of inflation and interest rate have similar patterns as the responses to monetary policy. It is important to highlight that the difficulty to identify shocks with sign restrictions is not restricted only to the transmission of monetary policy shocks; in fact, there are other situations where shocks might get confounded when only sign restrictions are imposed. As an example, consider the situation where the econometrician is interested in identifying the impact of demand shocks on the labor market and to identify the shock of interest she imposes sign restrictions on output, prices and interest rate (all moving in the same direction) and leaves the response of the labor market variable unrestricted. In the context of the SW model, two supply shocks (technology and price mark up shock) in combination can generate the same pattern of responses for output, inflation and interest rate as the demand shock (risk premium). In fact, a positive technology shock pushes output up and modestly but persistently inflation down. At the same time a markup shock increases inflation significantly in the short run but have little impact on output. The resulting net effect of these two shocks is an increase in both prices and quantities. Monetary policy responds to the shocks by rising rates and overall the impact of these supply shocks resemble a demand type disturbance. However, while the risk premium shock depresses the demand for hours worked, the impact of the mongrel shock (technology + price mark up shock) on hours worked is actually

[10] Our implementation of the Smets-Wouters model is based on Dynare (see Adjemian, Bastani, Juillard, Karamé, Maih, Mihoubi, Perendia, Pfeifer, Ratto and Villemot (2011)) replication code kindly provided by Johannes Pfeifer. The code is available at https://sites.google.com/site/pfeiferecon/dynare.

positive. As a result, sign restrictions would not be able to isolate the impact of the demand shock to the labor market variable.[11]

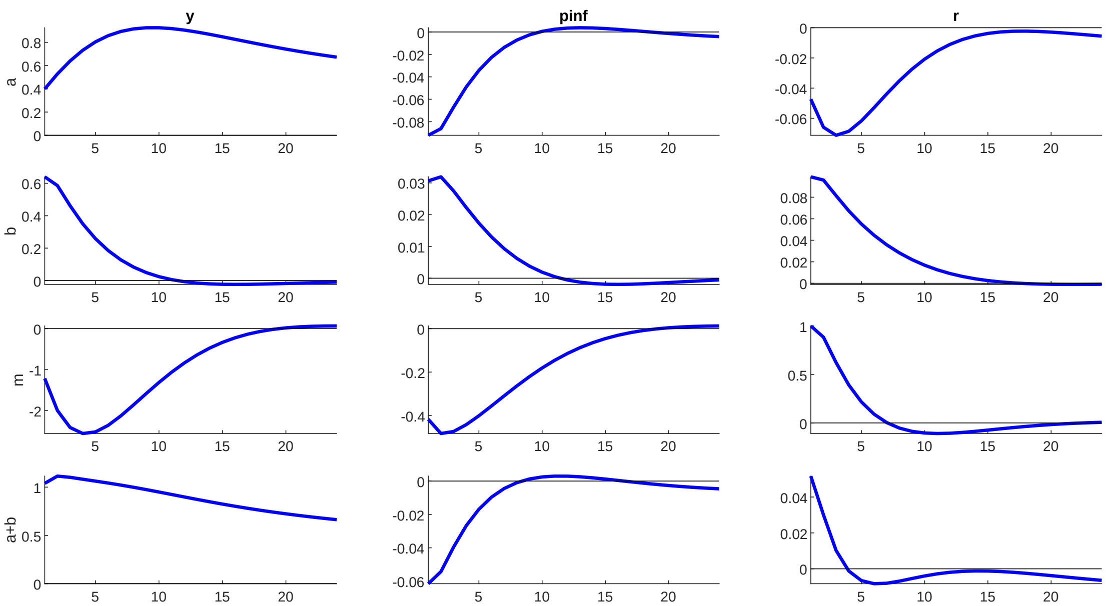


Figure 3: SW estimates of impulse response functions. From top to bottom technology, risk premium and monetary policy shocks and the sum of demand and supply shocks.

Simulated data We generate artificial data from the SW model by bootstrapping their estimated smoothed values for the technology, risk premium and monetary policy shocks. Figure 4 reports the estimated shocks realizations (top panels) and the empirical probability distribution against the normal one (bottom panels). Estimated shocks display deviations from normality. All shocks seems to display some form of fat-tails. In particular, we compute a robust measure of kurtosis using equation (1). Loosely speaking, this robust estimator considers the ratio between the distance of the percentiles in the tails and the distance between the percentiles close to the median. The larger the numerator, the thicker the tails; the smaller the denominator the more clustered is the distribution around the median and hence the more leptokurtic the shock is. The robust measure of excess kurtosis for the SW monetary policy shock equals 2, whereas any robust measure of skewness is very close to zero. Drawing randomly from the empirical distribution of the structural shocks, we generate a long sample of data consisting of 50,000 observations for output, inflation and interest rate, estimate the reduced form VAR parameters and study how monetary policy shocks transmit in the VAR with different set of restrictions and identification schemes.

[11]See appendix for more details.

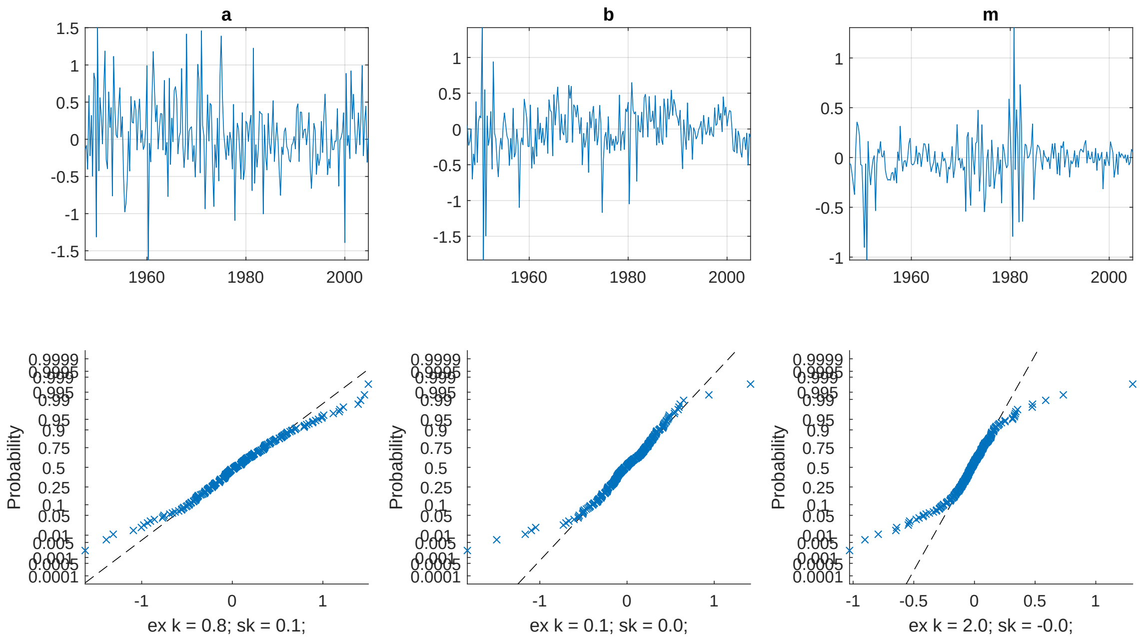


Figure 4: SW estimated shocks: form left to right technology, risk premium and monetary policy shocks. Top panels realizations, bottom panels probability distribution against the normal.

Identifying monetary policy shocks – large samples We study the identification of monetary policy shocks using sign restrictions and sign and higher-order moments restrictions. In particular, we assume that after a monetary policy shock,

- Inflation decreases on impact and for two consecutive quarters
- Interest rate increases on impact and for two consecutive quarters
- Monetary policy shocks are leptokrutic, i.e. monetary policy robust measure of excess kurtosis larger than 1.6

The threshold value for the monetary policy robust kurtosis is chosen to be larger than roughly 80% of the value estimated in the empirical distribution.[12] No restriction is imposed on output. We estimate a VAR with twenty lags and rotate the reduced-form Least Square VAR innovations so that the sing restrictions on the impulse responses are satisfied. We construct 50,000 accepted rotations using Rubio-Ramírez et al. (2010) algorithm. For each of these rotations, we verify if the higher moment restriction is verified; and if so, we keep the candidate rotation. We also compute the rotation matrix by using the eigenvalue decomposition discussed in Section 2. For

[12]Thresholds lower than one do not alter the identified set, see Figure 14 in the appendix, and the probability that output is positive one year after the tightening is about 40% both with sign and with sign and higher-order moment restrictions. This occurs because both demand and supply are somehow leptokurtic; however, their tails are not fat enough to masquerade as a monetary policy shock with leptokurticity larger than 1.3. with such a threshold we start seeing the HOM identified set to shrink relative to the sign restricted identified set.

each identification scheme we compute the impulse responses which are reported in figure 5, where the first raw reports the 90% and 99% identified sets of impulse responses using sign restrictions only and the second raw using sign and the higher moment inequality restriction on the monetary policy shock fourth moment. The blue line in each panel displays the true monetary policy impulse responses and the red dashed line the impulse responses computed using the eigenvalue decomposition of the matrix collecting all the fourth moments of the VAR empirical innovations.

As in the simple NK model case, the set of responses identified by sign restrictions is so wide that the impact of monetary policy shock on output is inconclusive. On the contrary, higher order moments help refining the transmission of monetary policy shocks to output. In the case of inequality restrictions, the 90% identified set suggests the most of the output trajectories are negative at least after few quarters in line with what the theory predicts. Interestingly, the identification using the spectral or eigenvalue decomposition of the empirical innovation fourth moment matrix singles out the correct rotation matrix and identifies very precisely the transmission of a monetary policy impulse. It is important to stress however that the quality of this identification rests on the fact that we have a long span of data and the estimates of the fourth order moments of the empirical innovations are very precise. With shorter samples, the quality of point identification deteriorates and inequality restrictions perform better as we show in the next section.

Finally it is important to notice that higher moments restrictions shrink the identified set also when we impose the sign restrictions that a monetary policy shock generates on the observed variables; figure 13 in the appendix report the magnitude of the refinement of the identified set in the context of the SW model.

Identifying monetary policy shocks – short samples The identification based on the spectral decomposition requires computing all the fourth moments of the vector of the VAR empirical innovations, a number which grows with the size of the VAR. Moreover, when the time series is short, sample counterparts of the fourth moments can be poorly estimated and be sensitive to outliers; thus the quality of point identification deteriorates significantly. In this respect, identification based on higher moment inequality restrictions has the advantage that it requires computing only the fourth moment of the shock of interest, which in our case is the kurtosis of the monetary policy shock. Robust measures of kurtosis can be easily computed and all perform well in small sample.

To argue that the latter argument is true, we run the following Montecarlo exercise. We constructed 500 different set of data each consisting of 200 observations length on output, inflation and interest rate. For each dataset we estimated a VAR and computed the monetary policy impulse responses using the spectral or eigenvalue decomposition and the sign and higher moment inequality restrictions. Figure 6 reports the dispersion of the (median) point estimate across identification schemes. With short samples, the spectral decomposition generates very dispersed point estimates which includes positive responses of output and prices after a monetary policy shock. This does not occurs with our preferred identification scheme, where point estimate of the responses of output and inflation are consistently estimated to be negative.

Finally, we also have computed the median value of the upper and lower bounds of the 68% high probability density (HPD) sets across different artificial samples. Results confirm that bounds of the HPD set response of output are negative with the sign and higher moment inequality restrictions. This is not the case for the eigenvalue decomposition nor the pure sign restrictions identification, figure 12.

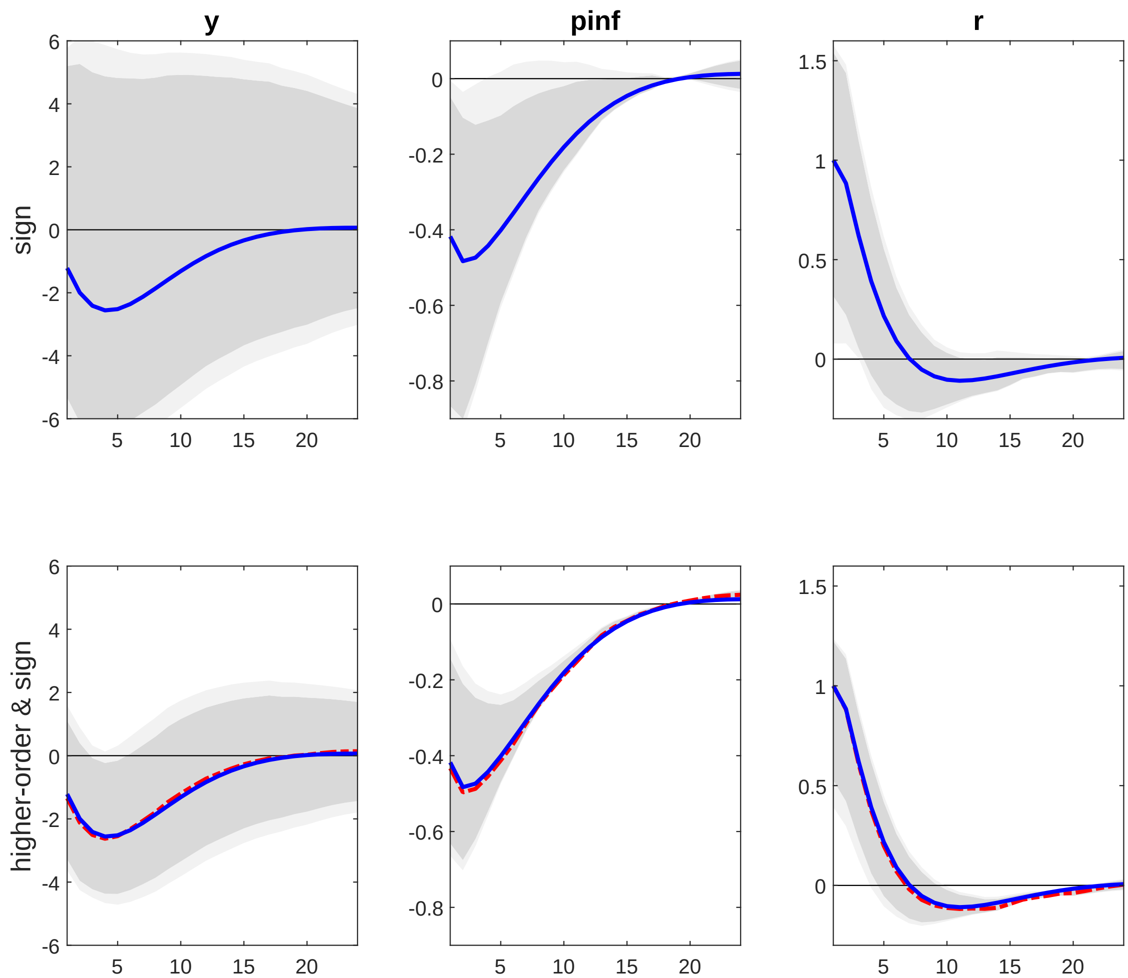


Figure 5: IRF using sign (first row) and sing and higher moment inequality (second row) restrictions. The blue solid line is the true impulse response. The red dashed line indicate the point identification using the eigenvalue decomposition of the fourth moments of the reduced-form Least Square VAR innovations. The dark (light) gray areas report the 90% (99%) identified set using inequality restrictions on the fourth moment of the monetary policy shock.

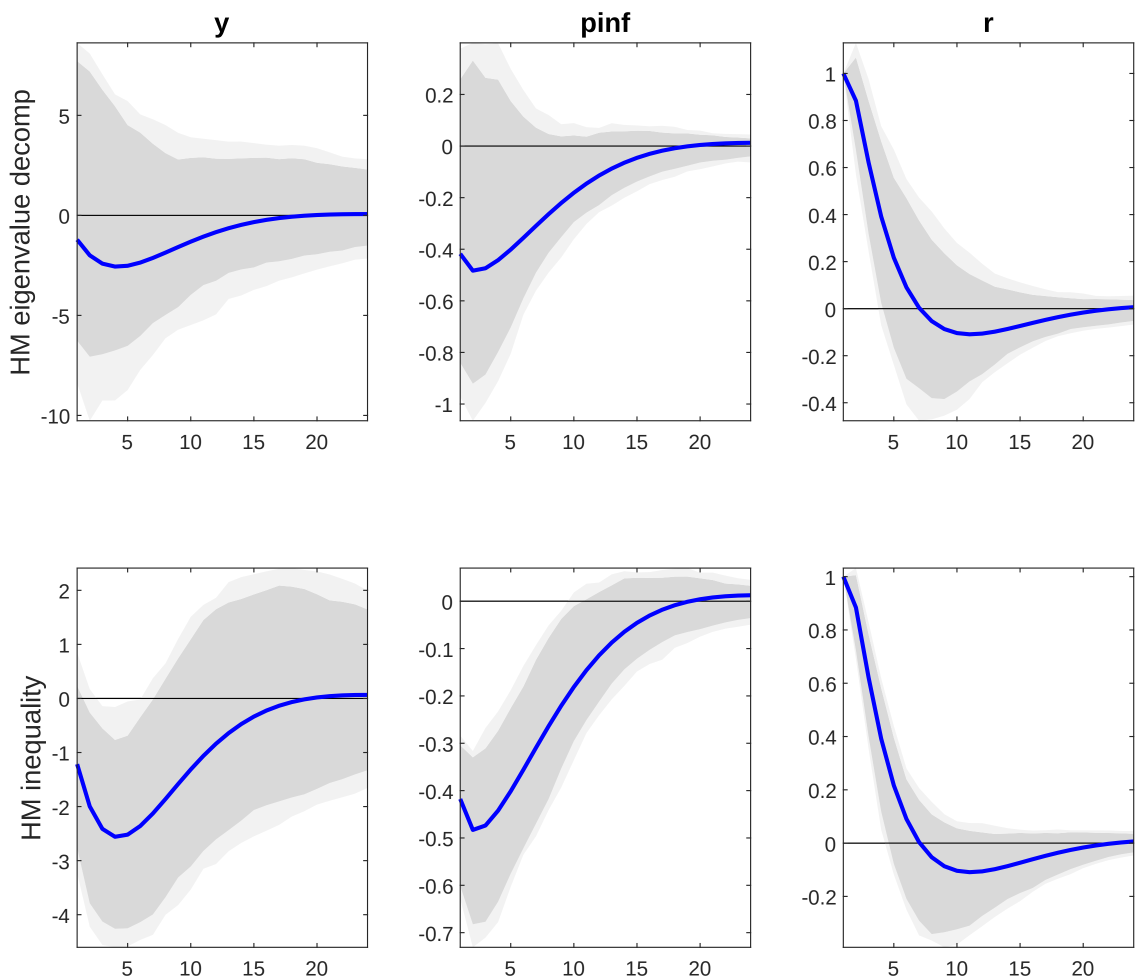


Figure 6: IRF using fourth moments eigenvalue decomposition and inequality restrictions (median IRF across samples). The dark (light) gray areas report the 90% (99%) dispersion of the point estimates over repeated samples of 300 observation length. The blue line is the true impulse response.

4 Empirical Application I: monetary policy in the US

Our first empirical application studies the transmission mechanism of conventional monetary policy in the US using a minimal set of assumptions about the sign pattern of the impulse responses in the same spirit of Uhlig (2005). The idea is to study the real effect of monetary policy without imposing any assumption on the reaction of output. The sign restrictions are complemented with the requirement that monetary policy shocks are drawn from a leptokurtic distributions, where most realizations are tiny but large deviations are more likely than with normal distributed shock. Next section discusses whether this assumption is reasonable.

4.1 Higher moments of monetary policy shock proxies

While structural shocks are not observable, various scholars have constructed measurable proxies for a number of structural shocks using a variety of methods and datasets, see Ramey (2016) for a survey. Scope of this section is to study the third and fourth moment properties of the estimated measures of monetary policy shocks, a particular type of demand shock for which we have a good number of measurements. One popular way to isolate monetary policy surprises is to look at the daily or intraday variations of interest rates around central bank monetary policy decisions and announcements. The high frequency of the data makes it more likely that such changes reflect unexpected shifts in the monetary policy stance. In particular, Jarociński (2021) showed that in a narrow window around FOMC announcements federal funds rate futures (and the SP500 index) variations are typically small but sometimes quite big suggesting that monetary policy surprises are leptokurtic. We show in this section that this property extends to many other observed proxies of monetary policy shocks studied in the literature and it is not specific to the U.S. experience.

In the context of the U.S. monetary policy surprises extracted from high frequency datasets we consider the monetary policy proxy used in Gertler and Karadi (2015) (GK), in Miranda-Agrippino and Ricco (2021) (MAR) and in Jarociński and Karadi (2020) (JK). Some of these proxies control for the information/Delphic effect of monetary policy (as discussed in Campbell, Evans, Fisher and Justiniano (2012) or Nakamura and Steinsson (2018)). We also look at the raw (unrotated) first three principal components of the intraday variations of the interest rates term structure around FOMC announcements (i.e. USf1-USf3). With lower frequency data, e.g. months, it is more difficult to isolate monetary policy surprises as other non-policy shocks can materialize and some structure is needed. Various scholars have used different methods to identify monetary policy with low frequency data, and here we consider the most popular ones. In particular, we consider the Romer and Romer (2004) (RR)[13] narrative instrument, Sims and

[13]We consider here the series constructed in Wieland and Yang (2020) who extend the Romer-Romer (2004) monetary

Zha (2006) (SZ) monetary policy innovations estimated from a SVAR with regimes shifts and the monetary policy shock estimated in the Smets and Wouters (2007) (SW) DSGE model.

For the Euro Area we consider the monetary policy surprises constructed in Andrade and Ferroni (2021), where they look at the high-frequency variations of future OIS contracts around the ECB monetary policy decisions and press conference and distinguish between conventional (AF(target)) and forward guidance shocks (AF(FWG)) controlling for information/Delphic effects (AF(delphic)). We also look at the unrotated first three principal components of variations of the Euro Area interest rates futures around the ECB decision and communication constructed in Altavilla, Brugnolini, Gürkaynak, Motto and Ragusa (2019) (EAf1-EAf3).

For the UK, Gerko and Rey (2017) construct monetary policy surprises by calculating the change of three-month sterling future rates around the release of the minutes of the MPC minute (GR(minute)) and inflation report (GR(IR)). The monetary policy surprises constructed in Cesa-Bianchi, Thwaites and Vicondoa (2020) (CBTV) is based on the changes in three months Libor around monetary policy events using the methodology in Gürkaynak, Sack and Swanson (2005). Kaminska and Mumtaz (2022) extract monetary policy surprises from the high frequency variations of the full yield curve of UK government bonds (KM). Finally, Cloyne and Hurtgen (2016) employ the Romer-Romer identification approach to construct a measure of UK monetary policy shocks (CH). We treat all these measures as contaminated proxies of the UK monetary policy shocks.

Table 1 reports the robust measures of skewness and excess kurtosis along with the 95% confidence intervals obtained by bootstrapping the series. While there is no clear evidence about skewness of the monetary policy surprises, all the proxies have a statistically significant excess kurtosis. The point estimates are very dispersed. For the US, the excess kurtosis ranges from a low of 1.4 to a upper estimate of 11 when we use the monetary policy innovation of GK. For the EA, the lower bound is 1.3 and the upper is 3.4. Similar considerations for the UK monetary policy measures. It is important to highlight that these observed measures of monetary policy shocks are constructed using different methods, different datasets, different time spans and countries. Yet, regardless of that all proxies are leptokurtic.

Finally, it is important to highlight that these different measures spam relatively different information sets; e.g. in the US the HF measures of monetary policy surprises and the narrative instrument constructed by RR are barely correlated, about 0.2 see figure 11 in the appendix. To the extend that they measure the same underlying shock, it is not straightforward to assess which proxy better characterizes monetary policy shocks and should be used to instrument the reduced form residuals VAR for identification. On the contrary, all monetary policy shock proxies are robustly leptokurtic; and this property being common across measures can be leveraged to refine

policy shock series.

the identification of monetary policy shocks.

	Ex-Kurtosis	Skewness	Sample Size	Sample Coverage
SW	2.0 [0.4, 3.2]	-0.0 [-0.1, 0.1]	179	1960-2004
SZ	3.8 [1.8, 6.3]	0.0 [-0.0, 0.1]	518	1960-2003
RR	3.2 [1.8, 5.1]	0.0 [-0.1, 0.1]	468	1969-2007
GK	11.3 [5.9, 18.2]	-0.3 [-0.3, -0.2]	269	1990-2012
MAR	3.3 [1.2, 5.9]	-0.1 [-0.2, 0.0]	228	1991-2009
JK	8.8 [5.4, 15.8]	-0.1 [-0.2, -0.0]	323	1990-2016
USf1	1.4 [0.3, 3.5]	0.1 [-0.0, 0.2]	204	1994-2017
USf2	3.0 [1.4, 7.1]	0.1 [-0.0, 0.2]	204	1994-2017
USf3	1.9 [0.5, 4.4]	0.0 [-0.1, 0.1]	204	1994-2017
AF(target)	2.5 [0.6, 5.3]	-0.0 [-0.2, 0.1]	134	2004-2015
AF(delphic)	1.3 [0.2, 3.9]	-0.0 [-0.2, 0.1]	134	2004-2015
AF(FWG)	1.4 [0.2, 3.6]	0.0 [-0.1, 0.1]	134	2004-2015
EAf1	3.4 [1.4, 5.6]	-0.0 [-0.2, 0.1]	197	2002-2019
EAf2	1.5 [0.3, 3.9]	-0.1 [-0.2, 0.0]	197	2002-2019
EAf3	1.1 [0.2, 3.4]	0.0 [-0.1, 0.2]	197	2002-2019
CH	13 [5.9, 38]	0 [-0.1, 0.1]	348	1997-2015
GR(minutes)	2.5 [1.3, 4.9]	-0 [-0.2, 0.1]	211	1997-2015
GR(IR)	3.6 [2.8, 4.6]	-0.1 [-0.2, 0.0]	211	1997-2015
CBTV	3.4 [0.6, 7.5]	-0.1 [-0.2, 0.0]	212	1979-2007
KM	5.3 [2.3, 11.2]	-0.0 [-0.1, 0.1]	235	1997-2016

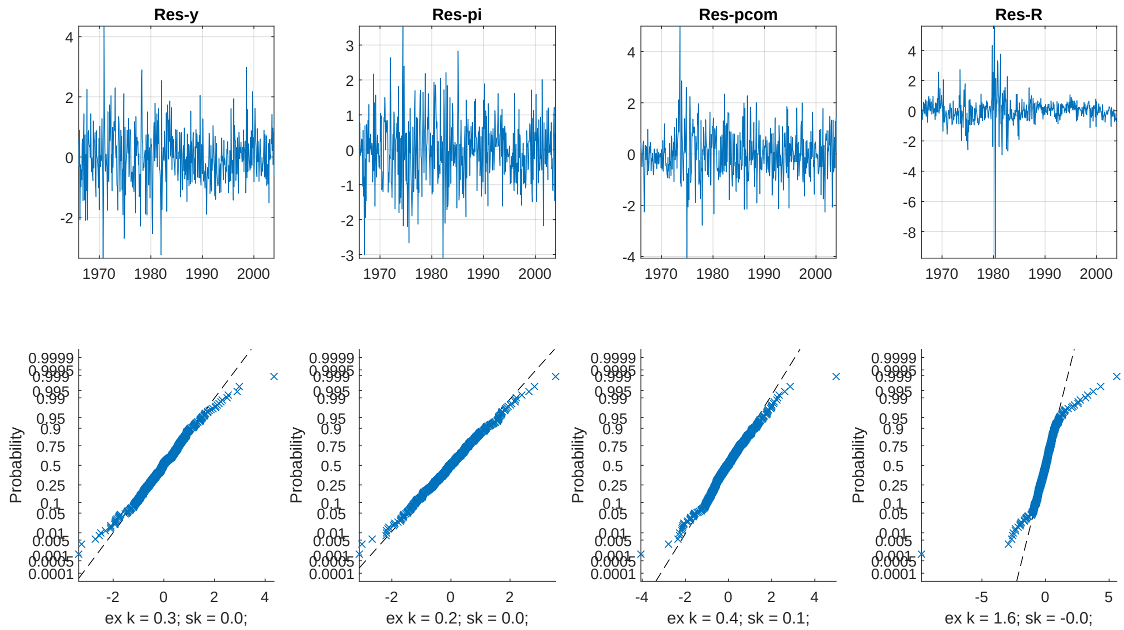
Table 1: Proxies of Monetary Policy shocks/surprises: Ex-Kurtosis & Skewness - Bootstrap; median and in parenthesis 95% confidence intervals. Proxies descriptions, sources and tags are in Table 2.

4.2 Monetary policy transmission with sign and kurtosis restrictions

We consider the dataset studied in Uhlig (2005) with observations on real activity, prices and interest rates from 1965m1 to 2003m12; in particular, the dataset consists of the Real GDP (y), the GDP Deflator (pi), the Commodity Price Index (pcom) and the Federal Funds Rate (FFR). We estimate the VAR parameters assuming 12 lags, obtain the whitened reduced form residuals and study their empirical probability distribution properties. As the least square (LS) estimates of the autoregressive coefficients are asymptotically consistent, so the LS residuals are treated as consistent estimators of the reduced form shocks.

Figure 7 plots the least square estimates of the whitened VAR reduced form shock, i.e. the reduced form residuals orthogonalized using the Cholesky factorization of the least square

Figure 7: Whitened Least Square residuals.



estimates of the residuals covariance matrix. The first row of the figure displays the estimated series and the second row compares the probability plot of the empirical distribution against a standard normal distribution (dash line); departures from the dash line indicate departures form the normality assumptions. Departures for normality are not fully evident for the output and prices whitened residuals; while there are few observations in the tails that lie far from the normal distribution implied ones, the Kolmogorov-Smirnov (K-S) test fails to reject the null that they could have come from a standard normal distribution; their respective $p-$values are 0.12 and 0.42. More evident departures from normality arise for the whitened residuals of the commodity prices and of the federal funds rate where the K-S test $p-$values are 0.07 and 0.00 respectively. The FFR residuals seems to be characterized by a large negative value at the beginning of the 80's. Even after removing the extreme values (min and max) of the FFR residuals, we still reject the null of normality with a $p-$value smaller than 1e-8. As a matter of fact the robust measure of skewness are all very close to zero, whereas the robust measures of kurtosis of the federal funds rate residual is significantly larger than those implied by the Gaussian distribution. Hence, the source of non-Gaussianity seems more evident in the tails thickness rather than in the asymmetry of the distribution.

Following the literature of the identification of monetary policy shocks using sign restrictions, we assume that monetary policy shocks move the interest rate and the price levels in opposite directions; in particular, we assume that an unexpected monetary policy shock induces an increase in the federal funds rate and a decline in the GDP deflator and in the price of commodities on impact and for the following six months. Based on our previous discussions and analysis, we assume further that the monetary policy shocks are significantly leptokurtic. In particular, we assume that the shock has the robust measure of excess kurtosis larger than 1.2. The choice of this value is motivated by the analysis in section 4.1 where the lowest value of the robust excess kurtosis across different estimates of US monetary policy surprises is 1.4, i.e. see the first column of table 1. As a reference, a random variable distributed –say– as mixture of two normal distributions where the random variable is drawn with a 0.8 (0.2) probability from (three times) the standard normal distribution has a robust excess kurtosis of 1.5. With this distribution, the probability of observing a realization distant two (three) standard deviations away from the mean is 25% (700 %) more likely that in the standard normal case.

Figure 8: Impulse responses to a monetary policy shock. Sign restrictions first row. Sign and kurtosis restrictions last row.

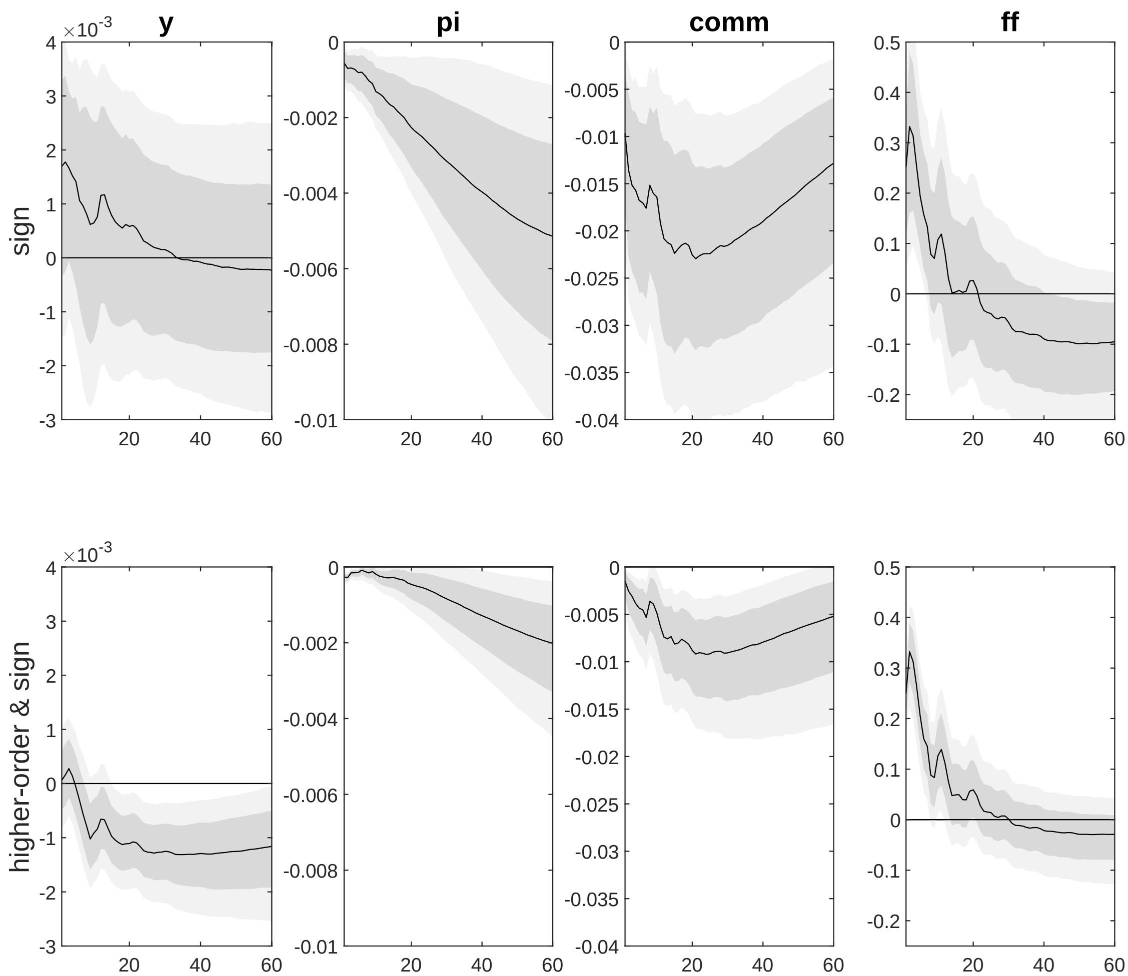


Figure 8 reports the impulse response function (IRF) of the variables of interest to a monetary policy shock. In the first row we reports the IRF identified with sign restrictions constructed using robust Bayesian methods and flat priors. The second row the analog IRF imposing sign and higher moment restrictions. In all panels the black line reports the median estimate and the (light) dark gray the 68 (90) % confidence sets. As discussed in section 3, monetary policy shocks have no clear effect on output when using only sign restrictions on interest rate and inflation are assumed. When we impose the inequality restriction on the excess kurtosis of monetary policy shocks, almost all the positive responses of output are chopped away and the distribution tilts towards negative outcomes only, as indicated by 68% and 90% bands in the bottom right panel of figure 8. So a monetary policy tightening consistently reduces output. This is not the only difference in the transmission of monetary policy shocks between the two identification schemes; magnitudes are also very different. For the same 25 bps increase in the federal fund rate on impact, prices decline significantly less, about half of the size, in the case where the inequality restriction on the monetary policy excess kurtosis are imposed, suggesting that the macroeconomic propagation of monetary policy is weaker when we isolate candidate shocks that display fat-tails.

5 Empirical Application II: spread shocks in the EA

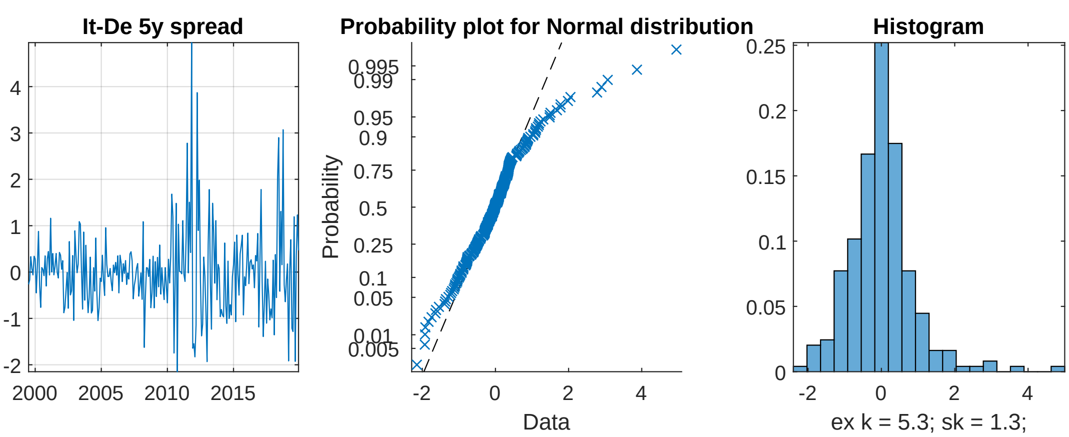
In the past decade the Euro Area has been characterized by large movements in sovereign spreads, i.e. the differential between the yield on long term government bonds of –say– Italy and Germany. While some of these movements are the results of changes in the economic fundamentals, some others are the results of political risks generating tensions in sovereign yield markets. Large movements in sovereign spreads have macroeconomic consequences but isolating the portion of these fluctuations due to changes in the economic fundamentals and those due to changes in political and sovereign risks is difficult. Some scholars have looked at financial market reactions around key political events, see e.g. Bahaj (2020) or Balduzzi, Brancati, Brianti and Schiantarelli (2023); the high frequency window makes it likely that the variations in market instruments are orthogonal to existing information and current economic conditions. Some other scholars have treated the sovereign spread itself as an exogenous process in structural dynamic general equilibrium models, see e.g. Bocola (2016) or Corsetti, Kuester, Meier and Muller (2013). In both setups, the default on sovereign debt evolves stochastically, capturing the uncertainty that surrounds the political process, and it is typically modeled as a random variable drawn from a Bernoulli distribution with time-varying probability; this often generates distributions for the endogenous variables with non-trivial third and fourth moments.[14] Hence,

[14]For example, Bocola (2016) studied the implications of an exogenous increase in sovereign risk for financial intermediation and showed that it can generate non-trivial third and fourth moments. In particular, the positive probability

higher moments restrictions can be thought in this context as characterizing sovereign risk or spread shocks and used for identification.

We considered EA data on industrial production (IP), core HICP (Core), unemployment rate, a measure of borrowing costs (EBP),[15] the one year Euribor, the spread between the 5 year Italian and German bond yield, and the 10 year Italian and German government bond yields from 1999m1 to 2019m12. We estimate a VAR model with six lags and uninformative priors for the parameters of interest. Using the LS estimates, we constructed first the LS reduced form residuals, and looked at their higher moments properties. In this case as well departures for normality are not evident for some variables. The K-S test rejects the null that the borrowing costs, the one year Euribor and the spread could have come from a standard normal distribution; for the remaining variables the test fails to reject the null. Figure 9 plots the least square estimates of the whitened VAR reduced form spread shock; the right panel of the figure displays the estimated series and the central panel compares the probability plot of the empirical distribution against a standard normal distribution (dash line), and the right panel shows the histogram. In this case, there is suggestive evidence that the spread is positively skewed and has fat-tails.

Figure 9: Whitened Least Square residual of the spread.



Our preferred identification scheme assumes that the spread shock increases the 5 year yield BTP-Bund spread, increase the 10 year Italian government bond yield and the borrowing costs on impact and for the following month; we assume further that the spread shock has sample skewness larger than one and sample excess kurtosis larger than three. For comparison, we also look at a pure sign restriction identification and at the recursive ordering identification where the spread shock is impacting macroeconomic and credit conditions only with a month

of a future default on the sovereign debt generates skewed and fat-tailed distributions of financial frictions which in his model are captured by the Lagrange multiplier on the incentive constraints of bankers.

[15]We use the non-financial corporation borrowing costs constructed in Gilchrist and Mojon (2017).

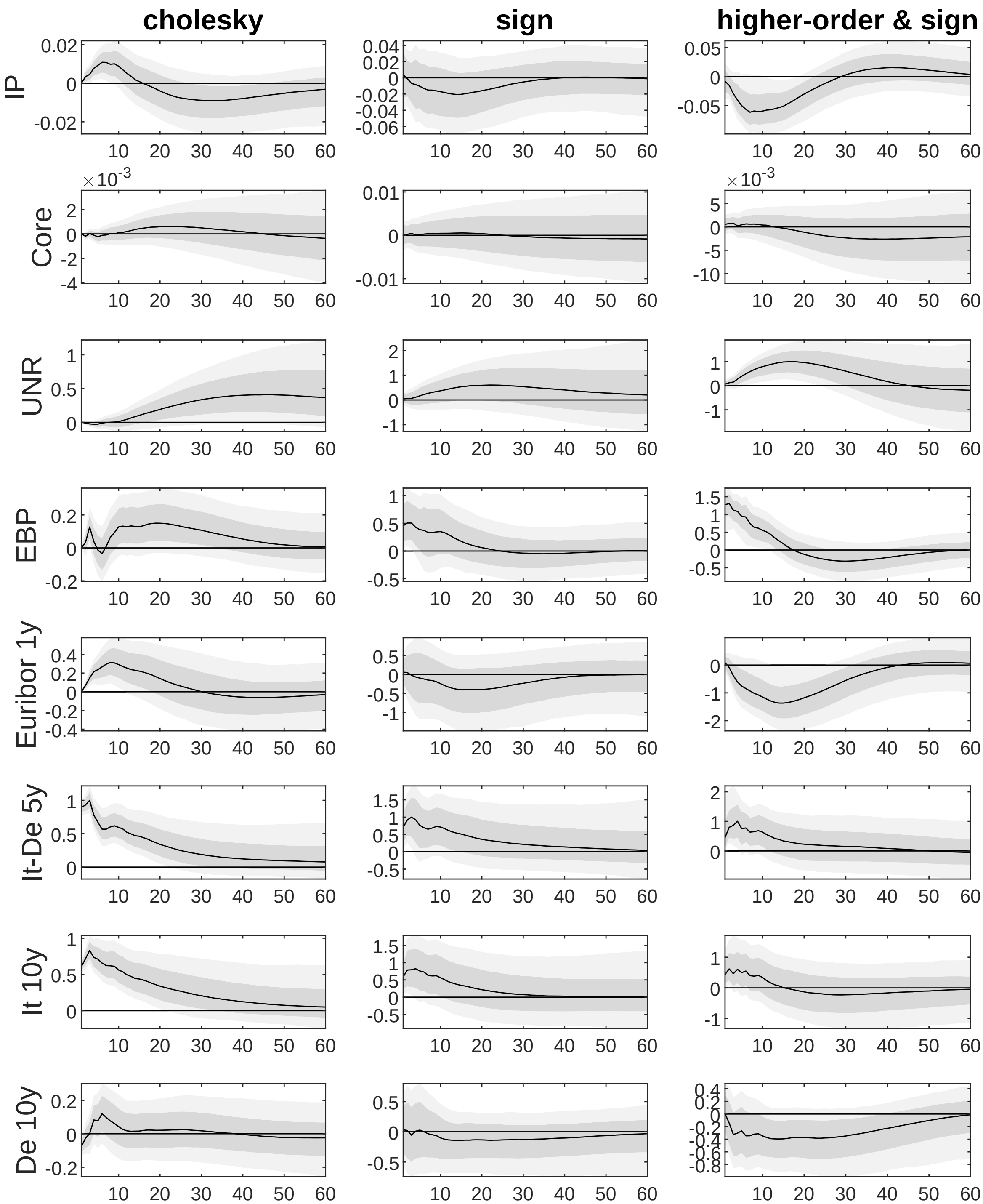
lag.[16] Figure 10 reports the impulse responses to a spread shock using a recursive ordering (first column), with sign restrictions (second raw), and sign and higher moment restrictions (third column); impulse responses are normalized so that the maximum median effect on the spread over the response horizon is one percent.

The identification using signs and higher moments restrictions produces interesting dynamic responses. An increase in spread of one percent generates an increase in the long term Italian bond of 50 bps and a flight to quality to Germany with a decline in the ten year yield of about 40 bps after one year. Credit conditions tighten immediately and borrowing costs increase more than one-to-one to the increase in the sovereign spread. Real variables also respond on impact after the spread shock: industrial production contracts and the unemployment rate increases. The peak effects after the shock are within six months for industrial production and eighteen months for unemployment. In less than three years the shock is absorbed. The price level for core good and services responds more sluggishly and slowly declines albeit non significantly. Overall these effects resemble those of a negative demand shock. In this depressed demand environment the one year Euribor drops suggesting an accomodative response of monetary policy. Most of these effects are significant statistically and in magnitudes, but short lasting.

The identification with a recursive ordering offers a very different picture, with some puzzling patterns. A one percent increase in the spread triggers an increase in the 10 year Italian bond yield, no impact on the German analog (if anything it increases) and an increase in the short term risk free rate. Credit costs tend to increase modestly, with the peak effect of the spread shock on borrowing costs occurring after 18 months at about one fifth of the size of the spread shocks. Industrial production increases in the short run and declines after two years; the unemployment rate increases for several months after the shock reaching its peak in three to four years after the impulse. The price level only modestly increases but it is not significant.

[16] The identification using only sign restrictions would not generate any statistically significant response.

Figure 10: Impulse responses to a spread shock. Recursive ordering first row. Sign restrictions second row. Sign, skewness and kurtosis restrictions last row. Impulse responses are normalized so that the maximum median impact on the spread is 1 percent.



6 Conclusions

We propose a novel set of necessary conditions based on higher moments to identify structural shocks from empirical innovations. Higher moment inequality restrictions constraint the structural shocks to have non-trivial third or fourth moments, i.e. different from the Gaussian-implied analogs. We show how the (sign-)identified set shrinks when these restrictions are introduced, both analytically and numerically. We show also that higher moment restrictions are economic relevant as they can sharpen the identification of economic shocks. In particular, we show how the excess kurtosis restriction can help isolating the impact of monetary policy shock on output from the supply and demand *masquerading* shock in standard New Keynesian (NK) models. Using a Bayesian robust approach we apply our identification scheme to study the transmission of conventional monetary policy shocks in the U.S. before the financial crisis and the propagation of spread shocks in the Euro Area.

References

Adjemian, S., Bastani, H., Juillard, M., Karamé, F., Maih, J., Mihoubi, F., Perendia, G., Pfeifer, J., Ratto, M. and Villemot, S.: 2011, Dynare: Reference manual version 4, *Dynare Working Papers 1*, CEPREMAP.

Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R. and Ragusa, G.: 2019, Measuring Euro Area Monetary Policy, *Journal of Monetary Economics* **108**, 162–179.

Andrade, P. and Ferroni, F.: 2021, Delphic and odyssean monetary policy shocks: Evidence from the euro area, *Journal of Monetary Economics* **117**, 816–832.

Antolin-Diaz, J. and Rubio-Ramírez, J. F.: 2018, Narrative Sign Restrictions for SVARs, *American Economic Review* **108**(10), 2802–2829.

Arias, J. E., Caldara, D. and Rubio-Ramírez, J. F.: 2019, The systematic component of monetary policy in svars: An agnostic identification procedure, *Journal of Monetary Economics* **101**, 1–13.

Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F.: 2018, Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications, *Econometrica* **86**, 685–720.

Bahaj, S.: 2020, Sovereign spreads in the euro area: Cross border transmission and macroeconomic implications, *Journal of Monetary Economics* **110**, 116–135.

Balduzzi, P., Brancati, E., Brianti, M. and Schiantarelli, F.: 2023, Political risk, populism, and the economy, *The Economic Journal* .

Bocola, L.: 2016, The pass-through of sovereign risk, *Journal of Political Economy* **124**(4), 879–926.

Bowley, A. L.: 1926, *Elements of statistics*, number 8, King.

Campbell, J. R., Evans, C. L., Fisher, J. D. and Justiniano, A.: 2012, Macroeconomic Effects of Federal Reserve Forward Guidance, *Brookings Papers on Economic Activity* (Spring), 1–80.

Cesa-Bianchi, A., Thwaites, G. and Vicondoa, A.: 2020, Monetary policy transmission in the united kingdom: A high frequency identification approach, *European Economic Review* **123**, 103375.

Christiano, L. J., Eichenbaum, M. and Evans, C. L.: 2005, Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy, *Journal of Political Economy* **113**(1), 1–45.

Cloyne, J. and Hurtgen, P.: 2016, The macroeconomic effects of monetary policy: A new measure for the united kingdom, *American Economic Journal: Macroeconomics* **8**(4), 75–102.

Comon, P.: 1994, Independent Component Analysis, a new concept?, *Signal Processing* **36**, 287–314.

Corsetti, G., Kuester, K., Meier, A. and Muller, G. J.: 2013, Sovereign risk, fiscal policy, and macroeconomic stability, *The Economic Journal* **123**(566), F99–F132.

Crow, E. L. and Siddiqui, M.: 1967, Robust estimation of location, *Journal of the American Statistical Association* **62**(318), 353–389.

Drautzburg, T. and Wright, J.: 2023, Refining Set-Identification in VARs through Independence, *Journal of Econometrics* **Forthcoming**.

Ferroni, F. and Canova, F.: 2021, A Hitchhiker's Guide to Empirical Macro Models, *Working Paper Series WP-2021-15*, Federal Reserve Bank of Chicago.

Galí, J.: 2015, *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications Second edition*, number 10495 in *Economics Books*, Princeton University Press.

Gerko, E. and Rey, H.: 2017, Monetary policy in the capitals of capital, *Journal of the European Economic Association* **15**(4), 721–745.

Gertler, M. and Karadi, P.: 2015, Monetary Policy Surprises, Credit Costs, and Economic Activity, *American Economic Journal: Macroeconomics* **7**(1), 44–76.

Gilchrist, S. and Mojon, B.: 2017, Credit Risk in the Euro Area, *Economic Journal* **128**, 118–158.

Gouriéroux, C., Monfort, A. and Renne, J.-P.: 2017, Statistical inference for independent component analysis: Application to structural var models, *Journal of Econometrics* **196**(1), 111–126.

Gouriéroux, C., Monfort, A. and Renne, J.-P.: 2019, Identification and Estimation in Non-Fundamental Structural VARMA Models, *The Review of Economic Studies* **87**(4), 1915–1953.

Groeneveld, R. A. and Meeden, G.: 1984, Measuring skewness and kurtosis, *Journal of the Royal Statistical Society: Series D (The Statistician)* **33**(4), 391–399.

Gürkaynak, R. S., Sack, B. and Swanson, E.: 2005, Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements, *International Journal of Central Banking* **1**(1), 55–93.

Hogg, R. V.: 1972, More light on the kurtosis and related statistics, *Journal of the American Statistical Association* **67**(338), 422–424.

Jarociński, M.: 2021, Estimating the Fed's unconventional policy shocks, *Working Paper Series 20210*, European Central Bank.

Jarociński, M. and Karadi, P.: 2020, Deconstructing Monetary Policy Surprises: The Role of Information Shocks, *American Economic Journal: Macroeconomics* **12**(2), 1–43.

Kaminska, I. and Mumtaz, H.: 2022, Monetary policy transmission during QE times: role of expectations and term premia channels, *Staff Working Paper 978*, Bank of England.

Kendall, M. G. and Stewart, A.: 1977, The advanced theory of statistics. vols. 1., *The advanced theory of statistics. Vols. 1.* **1**(Ed. 4).

Kilian, L. and Murphy, D. P.: 2012, Why Agnostic Sign Restrictions Are Not Enough: Understanding The Dynamics Of Oil Market Var Models, *Journal of the European Economic Association* **10**(5), 1166–1188.

Kim, T.-H. and White, H.: 2004, On more robust estimation of skewness and kurtosis, *Finance Research Letters* **1**(1), 56–73.

Kollo, T.: 2008, Multivariate skewness and kurtosis measures with an application in ica, *Journal of Multivariate Analysis* **99**(10), 2328–2338.

Lanne, M., Liu, K. and Luoto, J.: 2022, Identifying structural vector autoregression via leptokurtic economic shocks, *Journal of Business and Economic Statistics* .

Lanne, M., Meitz, M. and Saikkonen, P.: 2017, Identification and estimation of non-gaussian structural vector autoregressions, *Journal of Econometrics* **196**(2), 288–304.

Miranda-Agrippino, S. and Ricco, G.: 2021, The Transmission of Monetary Policy Shocks, *American Economic Journal: Macroeconomics* **13**(3), 74–107.

Moors, J.: 1988, A quantile alternative for kurtosis, *Journal of the Royal Statistical Society: Series D (The Statistician)* **37**(1), 25–32.

Nakamura, E. and Steinsson, J.: 2018, High Frequency Identification of Monetary Non-Neutrality, *Quarterly Journal of Economics* **133**(3), 1283–1330.

Petrova, K.: 2022, Asymptotically valid bayesian inference in the presence of distributional misspecification in var models, *Journal of Econometrics* **230**(1), 154–182.

Ramey, V. A.: 2016, Macroeconomic Shocks and Their Propagation, *in* J. B. Taylor and H. Uhlig (eds), *Handbook of Macroeconomics*, Vol. 2A, North Holland, chapter 3, pp. 71–162.

Romer, C. D. and Romer, D. H.: 2004, A New Measure of Monetary Shocks: Derivation and Implications, *American Economic Review* **94**(4), 1055–1084.

Rubio-Ramírez, J. F., Waggoner, D. F. and Zha, T.: 2010, Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference, *Review of Economic Studies* **77**(2), 665–696.

Schott, J.: 2016, *Matrix Analysis for Statistics*, Wiley Series in Probability and Statistics (Third Edition), Wiley.

Sims, C. A.: 1980, Macroeconomics and reality, *Econometrica* **48**(1), 1–48.

Sims, C. A. and Zha, T.: 2006, Were There Regime Switches in U.S. Monetary Policy?, *American Economic Review* **96**(1), 54–81.

Smets, F. and Wouters, R.: 2007, Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, *American Economic Review* **97(3)**, 586–606.

Uhlig, H.: 2005, What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**(2), 381–419.

Wieland, J. F. and Yang, M.-J.: 2020, Financial dampening, *Journal of Money, Credit and Banking* **52**(1), 79–113.

Wolf, C.: 2020, SVAR (Mis-)Identification and the Real Effects of Monetary Policy Shocks, *American Economic Journal: Macroeconomics* **12**(4), 1–32.

Wolf, C.: 2022, What Can We Learn From Sign-Restricted VARs?, *AEA Papers and Proceedings* **122**, 471–475.

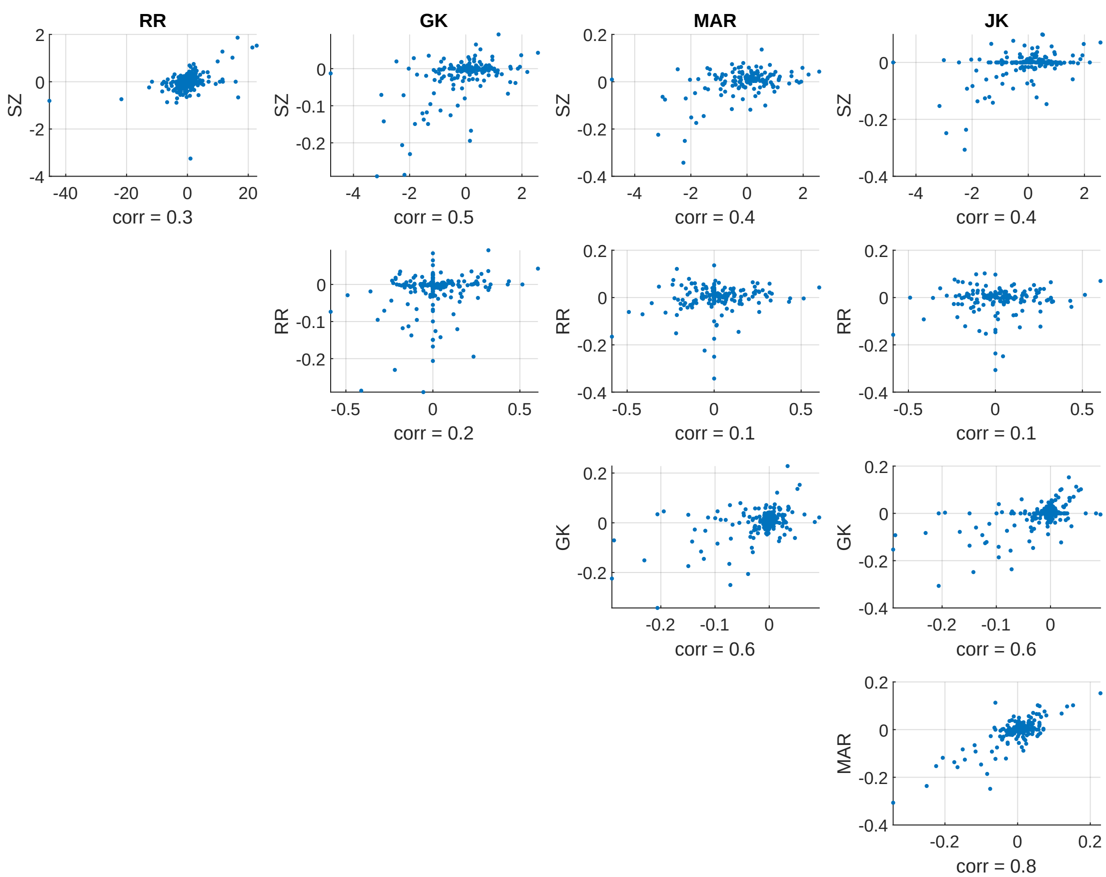
A Appendix

A.1 Additional Tables and Figures

tag	Paper	Description	Frequency	country
SZ	Sims and Zha (2006)	SVAR zero restrictions	M	US
RR	Romer and Romer (2004)	narrative	M	US
GK	Gertler and Karadi (2015)	HF	M	US
MAR	Miranda-Agrippino and Ricco (2021)	HF corrected for info-effect	M	US
JK	Jarociński and Karadi (2020)	HF corrected for info-effect	M	US
SW	Smets and Wouters (2007)	DSGE	Q	US
AF	Andrade and Ferroni (2021)	HF corrected for info-effect	M	EA
GK(M)	Gerko and Rey (2017)	HF around minutes	M	UK
GK(IR)	Gerko and Rey (2017)	HF around the inflation report	M	UK
CBTV	Cesa-Bianchi et al. (2020)	HF around monetary policy events	M	UK
KM	Kaminska and Mumtaz (2022)	HF around monetary policy events	M	UK
CH	Cloyne and Hurtgen (2016)	narrative	M	UK

Table 2: Various monetary policy surprises, estimates and sources.

Figure 11: Correlations across measures of U.S. monetary policy shocks.



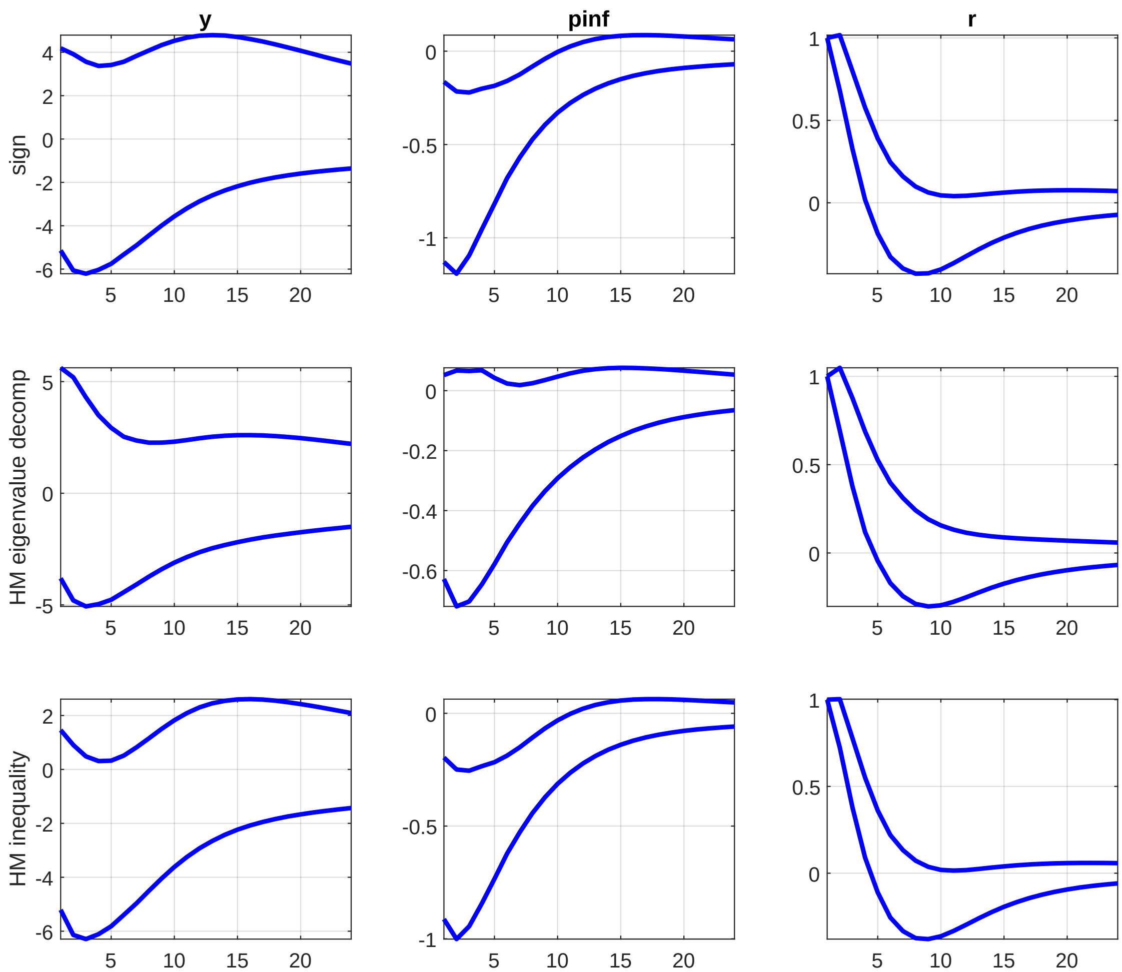


Figure 12: The upper and lower bounds of the 68% HPD set using fourth moments eigenvalue decomposition and inequality restrictions (median across samples).

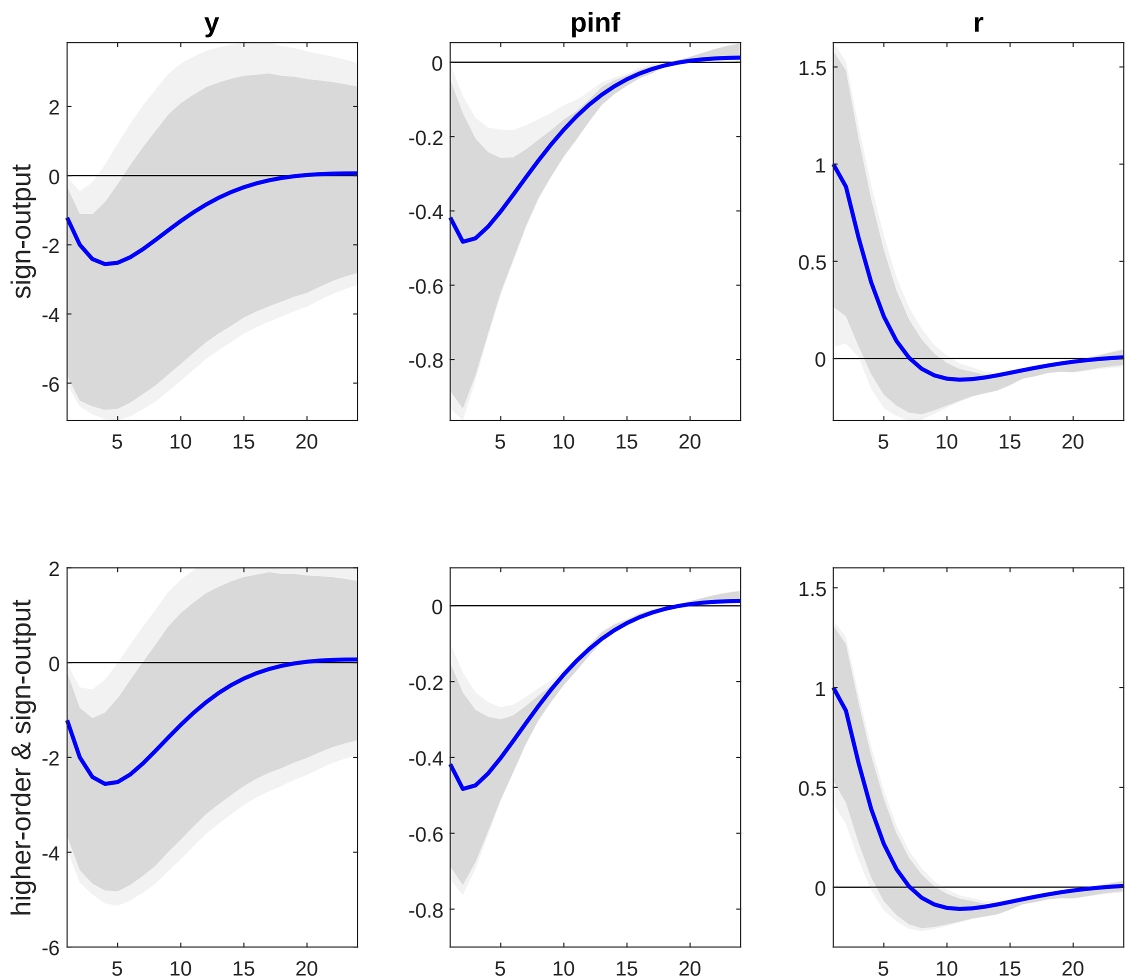


Figure 13: IRF using sign (first row) and sign and higher moment inequality (second row) restrictions (also on output). The blue solid line is the true impulse response. The dark (light) gray areas report the 90% (99%) identified set using inequality restrictions on the fourth moment of the monetary policy shock.

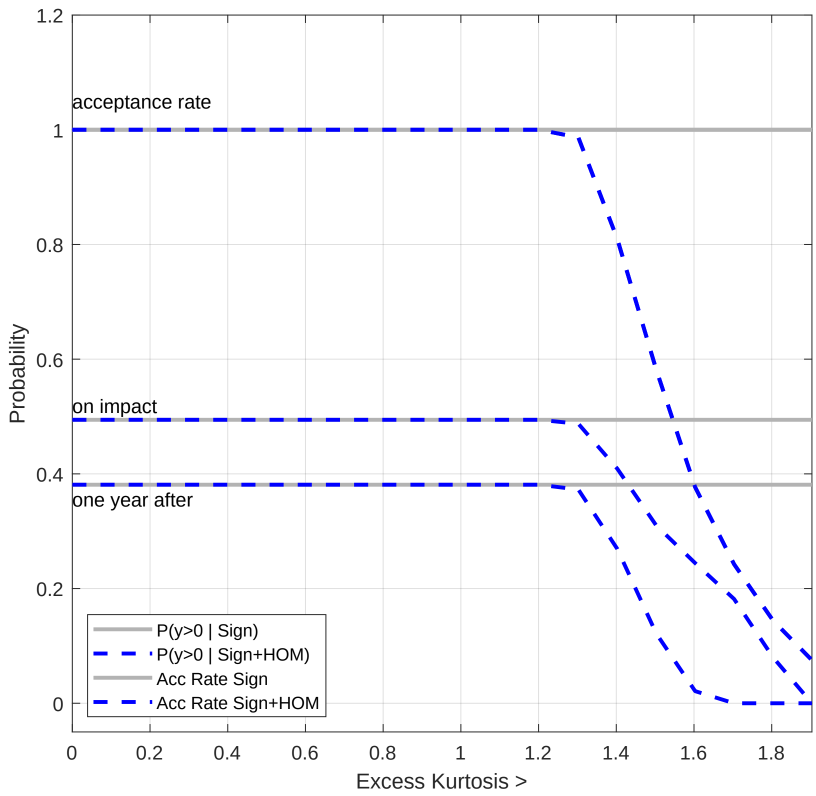


Figure 14: Probability of positive response of output after a monetary policy tightening as a function of the size of the interval.

A.2 Proof of Propositions

Notation: First, define $\boldsymbol{e}_k$ as the $n \times 1$ vector with zeros everywhere except a one in the k^{th} position, J_k the $n \times n$ matrix of zeros everywhere except one in the k^{th} position of the main diagonal and J_{jk} the $n \times n$ matrix of zeros everywhere except one in the $(j,k)^{th}$ element. For example, when $n = 3$, $k = 2$ and $j = 3$ we have

$$\boldsymbol{e}_2' = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, J_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } J_{3,2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Notice that

$$\begin{aligned}
\boldsymbol{e}_k' \otimes \boldsymbol{e}_k &= J_{k,k} = J_k \\
\boldsymbol{e}_j' \otimes \boldsymbol{e}_k &= J_{k,j} \\
\boldsymbol{e}_j' \otimes \boldsymbol{e}_k &= \boldsymbol{e}_k \boldsymbol{e}_j'
\end{aligned}$$

Denote with I_n the identity matrix of size n and with vec(X) the column-wise vecotrization of X. The following identities hold

$$\begin{aligned}
\text{vec}(I_n) &= \sum_{i=1}^{n} \boldsymbol{e}_i \otimes \boldsymbol{e}_i \\
\text{vec}(I_n)\text{vec}(I_n)' &= \left(\sum_{i=1}^{n} \boldsymbol{e}_i \otimes \boldsymbol{e}_i\right)\left(\sum_{i=1}^{n} \boldsymbol{e}_i \otimes \boldsymbol{e}_i\right)' = \left(\sum_{i=1}^{n} \boldsymbol{e}_i \otimes \boldsymbol{e}_i\right)\left(\sum_{i=1}^{n} \boldsymbol{e}_i' \otimes \boldsymbol{e}_i'\right) \\
&= \sum_{k,j=1}^{n} (\boldsymbol{e}_k \otimes \boldsymbol{e}_k)(\boldsymbol{e}_j' \otimes \boldsymbol{e}_j') = \sum_{k,j=1}^{n} \boldsymbol{e}_k \boldsymbol{e}_j' \otimes \boldsymbol{e}_k \boldsymbol{e}_j' = \\
&= \sum_{k,j=1}^{n} J_{k,j} \otimes J_{k,j}
\end{aligned}$$

Define the commutation matrix, $K_{n,n}$, the $(n^2 \times n^2)$ matrix consisting of $n \times n$ blocks where the $(j,i)-$element of the (i,j) block equals one, elsewhere there are all zeros. Notice that

$$K_{n,n} = \sum_{k,j=1}^{n} J_{k,j} \otimes J_{j,k}$$

Assumptions about the structural shocks ν: We assume that the strucutral shocks are independent and identically distributed over time. Moreover, we postulate that

- $E(\nu_{i,t}^2) = 1$ and $E(\nu_{i,t}\nu_{j,t}) = 0$ for all i, j
- $E(\nu_{i,t}^3) = \zeta_i$ and $E(\nu_{i,t}\nu_{j,t}\nu_{k,t}) = 0$ for all $i \neq j, k$
- $E(\nu_{i,t}^4) = \xi_i$, $E(\nu_{i,t}^2\nu_{j,t}^2) = 1$ for all $i \neq j$, and $E(\nu_{i,t}\nu_{j,t}\nu_{k,t}\nu_{m,t}) = 0$ for all $i \neq j, k, m$

Define the empirical innovation as

$$\iota_t = A_o \nu_t$$

where A_o is the true orthonormal rotation matrix. Finally, define the candidate structural shocks, $\breve{\nu}_t$, as $\breve{\nu}_t = A'\iota_t$, and denote with $\boldsymbol{\alpha}_k$ and $\boldsymbol{a}_k$ are the k^{th} column of A_o (the true impact matrix) and A (a candidate rotation) respectively.

Normal distribution fourth moments: In this section we show that the fourth moments of a multivariate normal distribution are invariant to orthonormal rotation matrix. Denote with $\mathcal{K}^z$ the matrix that collects the fourth moments of the *standard* normal distribution, which are given by

$$\mathcal{K}^z = I_{n^2} + K_{n,n} + \text{vec}(I_n)\text{vec}(I_n)'$$

see Kollo (2008) for more details. First, we show that $\mathcal{K}^z$ is invariant to orthonormal rotations. Using the property of the commutation matrix[17] that $K_{n,n}(A \otimes B) = (B \otimes A)K_{n,n}$, we have that the commutation matrix is invariant to any orthonormal rotation Ω, i.e. $(\Omega \otimes \Omega)' K_{n,n}(\Omega \otimes \Omega) = K_{n,n}(\Omega \otimes \Omega)'(\Omega \otimes \Omega) = K_{n,n}(\Omega' \otimes \Omega')(\Omega \otimes \Omega) = K_{n,n}(\Omega'\Omega \otimes \Omega'\Omega) = K_{n,n}$. Moreover, using the relationship between the vectorization and Kronecker product, i.e. $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, we have that $(\Omega \otimes \Omega)'(\text{vec}(I_n)\text{vec}(I_n)')(\Omega \otimes \Omega) = (\Omega' \otimes \Omega')\text{vec}(I_n)\ ((\Omega' \otimes \Omega')\text{vec}(I_n))' = \text{vec}(\Omega' I_n \Omega)\text{vec}(\Omega' I_n \Omega)' = \text{vec}(I_n)\text{vec}(I_n)'$. Therefore, we have that

$$(\Omega \otimes \Omega)'\ \mathcal{K}^z\ (\Omega \otimes \Omega) = \mathcal{K}^z$$

Denote with $\mathcal{K}^n$ the matrix that collects the fourth moments of the multivariate normal distribution with covariance Σ, which is given by

$$\mathcal{K}^n = (I_{n^2} + K_{n,n})(\Sigma \otimes \Sigma) + \text{vec}(\Sigma)\text{vec}(\Sigma)'$$

Using the same properties of matrices and Kronecker products it is straightforward to show that

$$\mathcal{K}^n = (\Sigma^{1/2} \otimes \Sigma^{1/2})'\ \mathcal{K}^z\ (\Sigma^{1/2} \otimes \Sigma^{1/2})$$

Third moments: Assume that $E(\nu_{n,t}^3) = \zeta_n \neq 0$. The $(n \times n^2)$ matrix collecting the structural shocks third moments can be written as

$$\begin{aligned}
E(\nu_t \nu_t' \otimes \nu_t') &= E(\nu_t \nu_t' \otimes [\begin{pmatrix}\nu_{1,t} & \dots & 0\end{pmatrix} + \dots + \begin{pmatrix}0 & \dots & \nu_{n,t}\end{pmatrix}]) = \\
&= E(\nu_{1,t}\nu_t\nu_t' \otimes \boldsymbol{e}_1' + \dots + \nu_{n,t}\nu_t\nu_t' \otimes \boldsymbol{e}_n') = \\
&= E(\nu_{1,t}\nu_t\nu_t') \otimes \boldsymbol{e}_1' + \dots + E(\nu_{n,t}\nu_t\nu_t') \otimes \boldsymbol{e}_n' = \\
&= \zeta_1 J_1 \otimes \boldsymbol{e}_1' + \dots + \zeta_n J_n \otimes \boldsymbol{e}_n' = \\
&= \sum_{k=1}^{n} \zeta_k J_k \otimes \boldsymbol{e}_k'
\end{aligned}$$

[17] See e.g. Schott (2016) (Theorem 8.26 at page 342)

Using the properties of the Kronecker product, i.e. $(A \otimes B)(C \otimes D) = AC \otimes BD$, notice that

$$\left(\sum_{i=1}^{n} \xi_i J_i \otimes \boldsymbol{e}_i\right)\left(\sum_{i=1}^{n} \xi_i J_i \otimes \boldsymbol{e}_i\right)' =$$

$$= \sum_{i=1}^{n} \zeta_i^2 \left(J_i \otimes \boldsymbol{e}_i'\right)\left(J_i \otimes \boldsymbol{e}_i'\right)' + \underbrace{\sum_{i \neq k}^{n} \zeta_i \zeta_k \left(J_i \otimes \boldsymbol{e}_i'\right)\left(J_k \otimes \boldsymbol{e}_k'\right)'}_{=0} =$$

$$= \sum_{i=1}^{n} \zeta_i^2 \left(J_i \otimes \boldsymbol{e}_i'\right)\left(\boldsymbol{e}_i \otimes J_i'\right) = \sum_{i=1}^{n} \zeta_i^2 \left(J_i \otimes \boldsymbol{e}_i'\right)\left(\boldsymbol{e}_i \otimes J_i'\right)$$

$$= \sum_{i=1}^{n} \zeta_i^2 J_i \boldsymbol{e}_i \otimes (J_i \boldsymbol{e}_i)' = \sum_{i=1}^{n} \zeta_i^2 \boldsymbol{e}_i \otimes \boldsymbol{e}_i'$$

$$= \sum_{i=1}^{n} \zeta_i^2 J_i' = \begin{pmatrix} \zeta_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \zeta_m^2 \end{pmatrix} = \Lambda_\zeta$$

where Λ_ζ is a diagonal matrix collecting the squared third moments of the structural shocks. Notice that the cross product are zero since $(J_i \otimes \boldsymbol{e}_i')(\boldsymbol{e}_k \otimes J_k') = J_k \boldsymbol{e}_i \otimes \boldsymbol{e}_k' J_i' = 0$.

The third moments of the candidate structural shocks are given by[18]

$$\begin{aligned}
E(\breve{\nu}_t \breve{\nu}_t' \otimes \breve{\nu}_t') &= A' E(\iota_t \iota_t' A \otimes \iota_t' A) = A' E(\iota_t \iota_t' \otimes \iota_t')(A \otimes A) = \\
&= A' E(A_o \nu_t \nu_t' A_o' \otimes \nu_t' A_o')(A \otimes A) = A' A_o E(\nu_t \nu_t \otimes \nu_t')(A_o' \otimes A_o')(A \otimes A) = \\
&= A' A_o E(\nu_t \nu_t \otimes \nu_t')(A_o' A \otimes A_o' A) = \\
&= A' A_o [\zeta_1 J_1 \otimes \boldsymbol{e}_1' + ... + \zeta_n J_n \otimes \boldsymbol{e}_n'](A_o' A \otimes A_o' A) = \\
&= \zeta_1 (A' A_o J_1 \otimes \boldsymbol{e}_1')(A_o' A \otimes A_o' A) + ... + \zeta_n (A' A_o J_n \otimes \boldsymbol{e}_n')(A_o' A \otimes A_o' A) = \\
&= \zeta_1 A' A_o J_1 A_o' A \otimes \boldsymbol{e}_1' A_o' A + ... + \zeta_n A' A_o J_n A_o' A \otimes \boldsymbol{e}_n' A_o' A
\end{aligned}$$

Notice that for all $k = 1, ..., n$,

$$\begin{aligned}
\boldsymbol{e}_k' A_o' A &= \begin{pmatrix} \boldsymbol{\alpha}_k' \boldsymbol{a}_1 & \boldsymbol{\alpha}_k' \boldsymbol{a}_2 & \dots & \boldsymbol{\alpha}_k' \boldsymbol{a}_n \end{pmatrix} \\
A' A_o J_k A_o' A &= \begin{pmatrix} \boldsymbol{a}_1' \boldsymbol{\alpha}_1 & \dots & \boldsymbol{a}_1' \boldsymbol{\alpha}_n \\ & \ddots & \\ \boldsymbol{a}_n' \boldsymbol{\alpha}_1 & \dots & \boldsymbol{a}_n' \boldsymbol{\alpha}_n \end{pmatrix} \begin{pmatrix} 0 & \dots & 0 \\ & 1 & \\ 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1' \boldsymbol{a}_1 & \dots & \boldsymbol{\alpha}_1' \boldsymbol{a}_n \\ & \ddots & \\ \boldsymbol{\alpha}_n' \boldsymbol{a}_1 & \dots & \boldsymbol{\alpha}_n' \boldsymbol{a}_n \end{pmatrix} = \\
&= \begin{pmatrix} (\boldsymbol{a}_1' \boldsymbol{\alpha}_k)(\boldsymbol{\alpha}_k' \boldsymbol{a}_1) & \dots & (\boldsymbol{a}_k' \boldsymbol{\alpha}_k)(\boldsymbol{\alpha}_k' \boldsymbol{a}_1) \\ & \ddots & \\ (\boldsymbol{a}_1' \boldsymbol{\alpha}_k)(\boldsymbol{\alpha}_k' \boldsymbol{a}_n) & \dots & (\boldsymbol{a}_n' \boldsymbol{\alpha}_k)(\boldsymbol{\alpha}_k' \boldsymbol{a}_n) \end{pmatrix} \\
&= A' \boldsymbol{\alpha}_k \boldsymbol{\alpha}_k' A
\end{aligned}$$

[18] We use the property of the kroeneker products such that $(A \otimes B)(C \otimes D) = AC \otimes BD$

Therefore we have for all $k = 1, ..., n$,

$$
\begin{aligned}
E(\breve{\nu}_t\breve{\nu}_t' \otimes \breve{\nu}_t') &= \sum_k \zeta_k A'A_oJ_kA_o'A \otimes \boldsymbol{e}_k'A_o'A = \\
&= \sum_k \zeta_k \begin{pmatrix} \boldsymbol{\alpha}_k'\boldsymbol{a}_1 \times A'A_oJ_kA_o'A & \boldsymbol{\alpha}_k'\boldsymbol{a}_2 \times A'A_oJ_kA_o'A & \dots & \boldsymbol{\alpha}_k'\boldsymbol{a}_n \times A'A_oJ_kA_o'A \end{pmatrix} \\
&= \begin{pmatrix} \sum_k \zeta_k\boldsymbol{\alpha}_k'\boldsymbol{a}_1 \times A'A_oJ_kA_o'A & \sum_k \zeta_k\boldsymbol{\alpha}_k'\boldsymbol{a}_2 \times A'A_oJ_kA_o'A & \dots & \sum_k \zeta_k\boldsymbol{\alpha}_k'\boldsymbol{a}_n \times A'A_oJ_kA_o'A \end{pmatrix} \\
&= \begin{pmatrix} \Phi^{(1)} & \Phi^{(2)} & \dots & \Phi^{(n)} \end{pmatrix}
\end{aligned}
$$

$E(\breve{\nu}_{n,t}^3)$ occupies the (n, n^2) position in the matrix $E(\breve{\nu}_t\breve{\nu}_t'\otimes\breve{\nu}_t')$, so we can focus on the $(n, n)-$element of $\Phi^{(m)}$, i.e.

$$
\begin{aligned}
\Phi^{(n)} &= \sum_k \zeta_k\boldsymbol{\alpha}_k'\boldsymbol{a}_n \times A'A_oJ_kA_o'A \\
&= \zeta_1\boldsymbol{\alpha}_1'\boldsymbol{a}_n \times A'A_oJ_1A_o'A + \dots + \zeta_n\boldsymbol{\alpha}_n'\boldsymbol{a}_n \times A'A_oJ_nA_o'A
\end{aligned}
$$

If $\boldsymbol{a}_n = \boldsymbol{\alpha}_n$, then $\boldsymbol{\alpha}_j'\boldsymbol{a}_n = 0$ if $j \neq n$ and $\boldsymbol{\alpha}_n'\boldsymbol{a}_n = 1$. Hence,

$$\Phi^{(n)} = \zeta_nA'A_oJ_nA_o'A$$

The $(n, n)-$element of $\Phi^{(n)}$ equals $\zeta_m(\boldsymbol{a}_n'\boldsymbol{\alpha}_n)(\boldsymbol{\alpha}_n'\boldsymbol{a}_n) = \zeta_m$. Finally, notice that if $\zeta_n = 0$, the impact column vector $\boldsymbol{a}_n$ of the matrix A is not identified. F

Fourth moments: Assume that $E(\nu_{n,t}^4) = \xi_n \neq 3$. The $(n^2 \times n^2)$ matrix collecting the full set of structural shocks fourth moments can be written as

$$
\begin{aligned}
\mathcal{K} &= E(\nu_t \nu_t' \otimes \nu_t' \otimes \nu_t) = E\left((\nu_{1,t}\nu_t\nu_t' \otimes \boldsymbol{e}_1' + ... + \nu_{n,t}\nu_t\nu_t' \otimes \boldsymbol{e}_n') \otimes \begin{pmatrix} \nu_{1,t} \\ \vdots \\ \nu_{n,t} \end{pmatrix} \right) = \\
&= E\left((\nu_{1,t}\nu_t\nu_t' \otimes \boldsymbol{e}_1' + ... + \nu_{n,t}\nu_t\nu_t' \otimes \boldsymbol{e}_n') \otimes (\nu_{1,t}\boldsymbol{e}_1 + ... + \nu_{n,t}\boldsymbol{e}_n) \right) = \\
&= E(\nu_{1,t}^2 \nu_t \nu_t') \otimes \boldsymbol{e}_1' \otimes \boldsymbol{e}_1 + ... + E(\nu_{n,t}^2 \nu_t \nu_t') \otimes \boldsymbol{e}_n' \otimes \boldsymbol{e}_n + \sum_{j \neq k} E(\nu_{j,t}\nu_{k,t}\nu_t\nu_t') \otimes \boldsymbol{e}_j' \otimes \boldsymbol{e}_k = \\
&= \begin{pmatrix} E\nu_{1,t}^4 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & E\nu_{1,t}^2 E\nu_{n,t}^2 \end{pmatrix} \otimes \boldsymbol{e}_1' \otimes \boldsymbol{e}_1 + ... + \begin{pmatrix} E\nu_{1,t}^2 E\nu_{n,t}^2 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & E\nu_{n,t}^4 \end{pmatrix} \otimes \boldsymbol{e}_n' \otimes \boldsymbol{e}_n \\
&\quad + \sum_{j \neq k} \begin{pmatrix} E(\nu_{1,t}^2 \nu_{j,t}\nu_{k,t}) & \cdots & E(\nu_{1,t}\nu_{n,t}\nu_{j,t}\nu_{k,t}) \\ & \ddots & \\ E(\nu_{1,t}\nu_{n,t}\nu_{j,t}\nu_{k,t}) & \cdots & E(\nu_{n,t}^2 \nu_{j,t}\nu_{k,t}) \end{pmatrix} \otimes J_{k,j} = \\
&= \begin{pmatrix} \xi_1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 1 \end{pmatrix} \otimes \boldsymbol{e}_1' \otimes \boldsymbol{e}_1 + ... + \begin{pmatrix} 1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \xi_n \end{pmatrix} \otimes \boldsymbol{e}_n' \otimes \boldsymbol{e}_n + \sum_{j \neq k} (J_{j,k} + J_{k,j}) \otimes J_{k,j} = \\
&= ((\xi_1 - 1)J_1 + I_n) \otimes J_1 + ... + ((\xi_n - 1)J_n + I_n) \otimes J_n + \sum_{j \neq k} (J_{j,k} + J_{k,j}) \otimes J_{k,j} = \\
&= \sum_{i=1}^{n} (\xi_i - 1) J_i \otimes J_i + I_n \otimes (J_1 + \cdots + J_n) + \sum_{j \neq k} (J_{j,k} + J_{k,j}) \otimes J_{k,j} = \\
&= \sum_{i=1}^{n} (\xi_i - 1) J_i \otimes J_i + I_{n^2} + \sum_{j \neq k \quad j,k=1}^{n} (J_{j,k} + J_{k,j}) \otimes J_{k,j}
\end{aligned}
$$

Since $\boldsymbol{e}_i' \otimes \boldsymbol{e}_i = J_i$ and $\boldsymbol{e}_j' \otimes \boldsymbol{e}_k = J_{k,j}$, for any i, j, k.

It is convenient to express the fourth moments in deviation from the standard normal distribution analogs. To this end define $\xi_i = x_i + 3$; when $x_i = 0$ then the fourth moments coincide with the standard normal distribution ones. We can rewrite the fourth moments of the structural shock as follows

$$
\begin{aligned}
\mathcal{K} &= \sum_{i=1}^{n}(x_i+2)J_i\otimes J_i + I_{n^2} + \sum_{j\neq k \quad j,k=1}^{n}(J_{j,k}+J_{k,j})\otimes J_{k,j} = \\
&= \sum_{i=1}^{n} x_i J_i\otimes J_i + I_{n^2} + 2\sum_{i=1}^{n} J_i\otimes J_i + \sum_{j\neq k \quad j,k=1}^{n}[J_{j,k}\otimes J_{k,j} + J_{k,j}\otimes J_{k,j}] = \\
&= \sum_{i=1}^{n} x_i J_i\otimes J_i + I_{n^2} + \left[\sum_{i=1}^{n} J_i\otimes J_i + \sum_{j\neq k \quad j,k=1}^{n} J_{j,k}\otimes J_{k,j}\right] + \left[\sum_{i=1}^{n} J_i\otimes J_i + \sum_{j\neq k \quad j,k=1}^{m} J_{k,j}\otimes J_{k,j}\right] = \\
&= \sum_{i=1}^{n} x_i J_i\otimes J_i + I_{n^2} + \sum_{j,k=1}^{m} J_{j,k}\otimes J_{k,j} + \sum_{j,k=1}^{n} J_{k,j}\otimes J_{k,j} = \\
&= \sum_{i=1}^{n} x_i J_i\otimes J_i + \underbrace{I_{n^2} + K_{n,n} + \mathrm{vec}(I_n)\mathrm{vec}(I_n)'}_{\mathcal{K}^z} \\
&= \sum_{i=1}^{n} x_i J_i\otimes J_i + \mathcal{K}^z
\end{aligned}
$$

where $K_{n,n}$ is the commutation matrix, vec is the column-wise vectorization of matrix and $\mathcal{K}^z$ denotes the matrix that collects the fourth moments of the standard normal distribution. It is straightforward now to derive the fourth moments of the candidate structural shocks

$$
\begin{aligned}
E(\breve{\nu}_t\breve{\nu}_t'\otimes\breve{\nu}_t'\otimes\breve{\nu}_t) &= E(A'\iota_t\iota_t'A\otimes\iota_t'A\otimes A'\iota_t) = E(A'[(\iota_t\iota_t'\otimes\iota_t')(A\otimes A)]\otimes A'\iota_t) = \\
&= E((A'\otimes A')[(\iota_t\iota_t'\otimes\iota_t')(A\otimes A)\otimes\iota_t]) = \\
&= E((A'\otimes A')[(\iota_t\iota_t'\otimes\iota_t')(A\otimes A)\otimes\iota_t\cdot 1]) = \\
&= E((A'\otimes A')[(\iota_t\iota_t'\otimes\iota_t'\otimes\iota_t)((A\otimes A)\otimes 1)]) = \\
&= (A'\otimes A')E(\iota_t\iota_t'\otimes\iota_t'\otimes\iota_t)(A\otimes A)
\end{aligned}
$$

With similar algebra we get

$$
\begin{aligned}
E(\breve{\nu}_t\breve{\nu}_t'\otimes\breve{\nu}_t'\otimes\breve{\nu}_t) &= (A'\otimes A')(A_o\otimes A_o)E(\nu_t\nu_t'\otimes\nu_t'\otimes\nu_t)(A_o'\otimes A_o')(A\otimes A) = \\
&= (A'\otimes A')(A_o\otimes A_o)\left(\sum_{i=1}^{n} x_i J_i\otimes J_i + \mathcal{K}^z\right)(A_o'\otimes A_o')(A\otimes A) = \\
&= \mathcal{K}^z + \underbrace{(A'\otimes A')(A_o\otimes A_o)\left(\sum_{i=1}^{n} x_i J_i\otimes J_i\right)(A_o'\otimes A_o')(A\otimes A)}_{\mathcal{K}^\star}
\end{aligned}
$$

In this case as well, with no departure from the normality (i.e. $x_i = 0$ for all i) the rotation matrix cannot be identified. Since we focus on the kurtosis of the shock ordered last, we are interested in the $(n^2, n^2)-$element of the matrix $E(\breve{\nu}_t\breve{\nu}_t'\otimes\breve{\nu}_t'\otimes\breve{\nu}_t)$ we can disregard the first matrix $\mathcal{K}^z$ and only consider the $(n^2, n^2)-$ element of the matrix $\mathcal{K}^\star$.

$$
\begin{aligned}
\mathcal{K}^\star &= (A' \otimes A')(A_o \otimes A_o)\left(\sum_{i=1}^{n} x_i J_i \otimes J_i\right)(A_o' \otimes A_o')(A \otimes A) = \\
&= (A'A_o \otimes A'A_o)\left(\sum_{i=1}^{n} x_i J_i \otimes J_i\right)(A_o'A \otimes A_o'A) = \\
&= \sum_{i=1}^{n} x_i (A'A_o J_i A_o' A) \otimes (A'A_o J_i A_o' A) = \\
&= \sum_{i=1}^{n} x_i (A'\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i' A) \otimes (A'\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i' A)
\end{aligned}
$$

The $(n^2, n^2)-$element of $\mathcal{K}^\star$ is the sum of the square of the $(n, n)-$elements of the matrices $A'\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'A$ times x_i for $i = 1, \ldots, n$, which is given by

$$
\begin{aligned}
\mathcal{K}^\star_{(m^2,m^2)} &= x_1(\boldsymbol{a}_n'\boldsymbol{\alpha}_1)^2(\boldsymbol{\alpha}_1'\boldsymbol{a}_n)^2 + \cdots + x_m(\boldsymbol{a}_n'\boldsymbol{\alpha}_n)^2(\boldsymbol{\alpha}_n'\boldsymbol{a}_n)^2 = \\
&= x_1(\boldsymbol{a}_n'\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1'\boldsymbol{a}_n)^2 + \cdots + x_m(\boldsymbol{a}_n'\boldsymbol{\alpha}_n\boldsymbol{\alpha}_n'\boldsymbol{a}_n)^2
\end{aligned}
$$

Hence we have that $\mathcal{K}^\star_{(n^2,n^2)} + \mathcal{K}^z_{(n^2,n^2)} = x_n + 3 = \xi_n$, if $\boldsymbol{a}_n = \boldsymbol{\alpha}_n$ (since $\boldsymbol{\alpha}_j'\boldsymbol{a}_n = 0$ if $j \neq n$ and $\boldsymbol{\alpha}_n'\boldsymbol{a}_n = 1$).

Alternative eigenvector decomposition of fourth moments The alternative eigenvector decomposition of fourth moments is based on Kollo (2008) and it requires more notation. The approach computes first the sum of the n^2 blocks of $(n \times n)$ sub-matrices of the fourth moment matrix and then take the eigenvalue/vector decomposition of the resulting matrix. The eigenvectors associated to non-zero eigenvalues coincide with the columns of the original rotation matrix up to a sign switch and permutation of columns.

Definition Let A be an $m \times n$ matrix and B an $mr \times ns$ partitioned matrix consisting of $r \times s$ blocks $B_{i,j}$, $i = 1, ..., m$ and $j = 1, ..., n$. The star product $A \star B$ of A and B is an $r \times s$ matrix such that

$$
A \star B = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j} B_{i,j}
$$

Define with $\mathcal{I}_n$ the $n \times n$ matrix of ones. Notice that

$$
\begin{aligned}
\mathcal{I}_n \star E(\iota_t \iota_t' \otimes \iota_t' \otimes \iota_t) &= \mathcal{I}_m \star (A_o \otimes A_o)(\mathcal{K} - \mathcal{K}_z)(A_o \otimes A_o)' \\
&= \sum_{i,j=1}^{n} (A_o \otimes \alpha_i)(\mathcal{K} - \mathcal{K}_z)(\alpha_j \otimes A_o)'
\end{aligned}
$$

We use of the Lemma 1 in Kollo (2008), which we report here

Lemma Let M, N, U and V be $n \times n$ matrices and a, b be $n \times 1$-vectors. Then

$$
(M \otimes a')diag(K_{n,n})(U \otimes V)diag(K_{n,n})(b \otimes N) = M diag(a)(U \circ V)diag(b)N
$$

where $\circ$ denotes the elementwise (Hadamard) product of matrices and *diag* is the diagonal matrix constructed on the vector.

Let Λ_ξ be the diagonal matrix collecting the excess fourth moments of the structural shocks, i.e. $E(s_{i,t}^4) - 3$; we have that

$$\mathcal{K} - \mathcal{K}_z = diag(K_{n,n})(\Lambda_\xi \otimes I_n)diag(K_{n,n})$$

We then have

$$\begin{aligned}
\mathcal{I}_m \star \Xi_4 &= \sum_{i,j=1}^{m} (A_o \otimes \alpha_i)(\mathcal{K} - \mathcal{K}_z)(\alpha_j \otimes A_o)' \\
&= \sum_{i,j=1}^{m} (A_o \otimes \alpha_i) diag(K_{n,n})(\Lambda_\xi \otimes I_n) diag(K_{n,n})(\alpha_j \otimes A_o)' \\
&= \sum_{i,j=1}^{m} A_o(\Lambda_\xi \circ I_n) diag(\alpha_i) diag(\alpha_j) A_o' \\
&= A_o \left(\sum_{i,j=1}^{m} (\Lambda_\xi \circ I_n) diag(\alpha_i) diag(\alpha_j) \right) A_o'
\end{aligned}$$

where the matrix in square brackets is diagonal. Therefore we can compute the eigenvalue/vetor decomposition of the $n \times n$ matrix $\mathcal{I}_m \star E(\iota_t \iota_t' \otimes \iota_t' \otimes \iota_t)$ and retrieve the impact matrix.

A.3 Examples

A.3.1 Bivariate case - third moments

To derive the identified set we use the following trigonometric identities:

$$a \cos x + b \sin x = \text{sgn}(a)\sqrt{a^2 + b^2} \cos\left(x + \arctan\left(-\frac{b}{a}\right)\right)$$

$$x = \arctan(\tan(x)); \quad \tan x = \frac{\sin x}{\cos x}; \quad \tan(-x) = -\tan x.$$

The higher-moment restrictions is then

$$\begin{aligned}
(\cos\theta \cos\theta_o + \sin\theta \sin\theta_o)^3 &> 0 \\
(\text{sgn}(\cos\theta_o)\cos(\theta - \theta_o))^3 &> 0
\end{aligned}$$

A.4 Robust estimates of kurtosis and skewness

Estimates of kurtosis and skewness based on fourth and third sample moments can be very sensitive to outliers with short samples, see Kim and White (2004). When shocks are distributed with t-student with low degrees of freedom, not even a million draws are enough to have median unbiased estimates of kurtosis, see Ferroni and Tracy (2022). A number of robust measures have been proposed in the literature and they all perform well for the sample size typically considered in macroeconomics. These robust estimators are constructed using ratios of distance between different percentiles of the empirical distribution. Loosely speaking, robust measures of kurtosis consider the ratio between the distance of the percentiles in the tails and the distance between the percentiles close to the median. The larger the numerator, the thicker the tails; the smaller the denominator the more clustered is the distribution around the median. Hence, with distributions centered around zero, realizations are often very small but sometime quite big. Robust measures of skewness exploit the distance between median and mean. More formally, let x be a random variable with cumulative density function F, the robust measures of kurtosis considered are:

- Moors (1988) kurtosis:
$$\mathcal{K}_m(x) = \frac{(p_7 - p_5) + (p_3 - p_1)}{p_2 - p_4}$$
where p_j represents the octile of the empirical distribution of x, i.e. $p_j = F^{-1}(j/8)$ with $j = 1, ..., 7$. If x follows a Gaussian distribution, then $\mathcal{K}_m(x)$ equals 1.23. Therefore the excess kurtosis is given by $\mathcal{EK}_m(x) = \mathcal{K}_m(x) - 1.23$.

- Hogg (1972) kurtosis:
$$\mathcal{K}_h(x) = \frac{u_{0.05} - l_{0.05}}{u_{0.5} - l_{0.5}}$$
where $u_\alpha(l_\alpha)$ is the average of the upper (lower) α percentile of the distribution of x. If x follows a Gaussian distribution, then $\mathcal{K}_h(x)$ equals 2.59. Therefore the excess kurtosis is given by $\mathcal{EK}_h(x) = \mathcal{K}_h(x) - 2.59$.

- Crow and Siddiqui (1967) kurtosis:
$$\mathcal{K}_{cs}(x) = \frac{F^{-1}(0.975) - F^{-1}(0.025)}{F^{-1}(0.75) - F^{-1}(0.25)}$$
where $F^{-1}(\alpha)$ is the α percentile of the distribution of x. If x follows a Gaussian distribution, then $\mathcal{K}_{cs}(x)$ equals 2.91. Therefore the excess kurtosis is given by $\mathcal{EK}(x)_{cs} = \mathcal{K}_{cs}(x) - 2.91$.

Robust measures of skewness considered are:

- Bowley (1926) skewness:
$$\mathcal{S}_b(x) = \frac{p_3 + p_1 - 2 \times p_2}{p_3 - p_1}$$

where p_j represents the quartiles of the empirical distribution of the candidate shock, i.e. $p_j = F^{-1}(j/4)$ with $j = 1, 2, 3$.

- Groeneveld and Meeden (1984) skewness:

$$\mathcal{S}_{gm}(x) = \frac{\text{mean} - \text{median}}{E|x - \text{median}|}$$

where $E|x - \text{median}|$ represents the average of the absolute deviation from the median.

- Kendall and Stewart (1977) skewness:

$$\mathcal{S}_{ks}(x) = \frac{\text{mean} - \text{median}}{\sigma}$$

where σ is the standard deviation of the empirical distribution of the candidate shock.

Clearly, all these robust measures of skewness are zero with the normal distribution.

The sets identified with excess kurtosis and skewness restriction on the shock $\nu_{1,t}$ are defined respectively as

$$\begin{aligned}\mathcal{A}_k &\equiv \{A \in \mathcal{O} | \; \mathcal{K}_i(\breve{s}_{1,t}) > k \text{ with } \breve{\nu}_t = A' \iota_t\} \\ \mathcal{A}_s &\equiv \{A \in \mathcal{O} | \; \mathcal{S}_j(\breve{s}_{1,t}) > k \text{ with } \breve{\nu}_t = A' \iota_t\} \end{aligned} \tag{2}$$

with i = Moore, Hogg or Crow-Siddiqui and with j = Bowley, Groeneveld-Meeden or Kendall-Stewart.

A.5 Estimation and Identification

In this section we briefly describe the estimation and identification strategy that we use to estimate a VAR with non-Gaussian errors and to construct IRF using higher-order moment restrictions. Let a $VAR(p)$ be:

$$y_t = \Phi_1 y_{t-1} + ... + \Phi_p y_{t-p} + \Phi_0 + u_t$$

where y_t is $n \times 1$ vector of endogenous variables, Φ_0 is a vector of constant and Φ_j are $n \times n$ matrices. We assume $y_0, \ldots, y_{-p+1}$ are fixed. We assume that u_t are i.i.d. zero mean random vectors with unconditional covariance matrix Σ. We assume that the VAR reduced form shocks are linear combination of the unobserved structural shocks, ν_t, i.e.

$$u_t = \Sigma^{1/2} \iota_t = \Sigma^{1/2} \; \Omega \; \nu_t$$

where $\Sigma^{1/2}$ is the Cholesky factorization of Σ and Ω is an orthonormal matrix, i.e. $\Omega\Omega' = \Omega'\Omega = I$. The structural shocks, ν_t, are zero-mean orthogonal shocks with unitary variance, i.e. $\nu_t \sim (0, I)$.

Standard inference on VAR parameters typically postulates a multivariate normal distribution for the reduced form innovations. Such an assumption cannot be considered in our context. We propose to adopt a robust Bayesian approach which allows to construct posterior credible sets without the need for distributional assumptions of the reduced form residuals. The Bayesian approach we use builds on the work by Petrova (2022), where she propose a robust and computationally fast Bayesian procedure to estimate the reduced form parameters of the VAR in the presence of non-Gaussianity. While Bayesian inference about the autoregressive coefficients is asymptotically unaffected by the distribution of the error terms, inference on the intercept and the covariance matrix are invalid in the presence of skewness and kurtosis.

The robust approach relies on the asymptotic normality of the Quasi Maximum Likelihood (QML) estimator[19] of reduced form parameters, autoregressive coefficients and covariances; Petrova (2022) derives the closed form expression for the asymptotic covariance matrix of the QML estimator allowing for fast simulation from its asymptotic distribution. In the case of symmetric distribution (no skewness), she shows that asymptotic valid inference for Σ can be performed by drawing from the asymptotic normal distribution centered in the consistent estimator of Σ, i.e. the QML estimator $\hat{\Sigma}$, and with covariance matrix equal to $\frac{1}{T}\left(\widehat{\mathcal{K}} - \text{vech}(\hat{\Sigma})\text{vech}(\hat{\Sigma})'\right)$, where $\widehat{\mathcal{K}}$ is a consistent estimator of the fourth moment of the VAR reduced form shocks. Importantly, as previously mentioned, sample counterparts of the fourth moments can be poorly estimated and be sensitive to outliers. We then follow Petrova (2022) and consider a shrinkage approach which consists in tilting the sample fourth moments estimates of the reduced form orthogonalized errors towards the normal distributed counterparts. In particular, the shrinkage estimator for the kurtosis is defined as

$$\widehat{\mathcal{K}}^{\star} = \frac{T}{T+\tau}\widehat{\mathcal{K}}_T + \frac{\tau}{T+\tau}D_n^{+}(I_n + K_{n,n} + \text{vec}(I_n)\text{vec}(I_n)')D_n^{+'} \tag{3}$$

where $\widehat{\mathcal{K}}_T$ represents the sample fourth moments of the orthogonalized reduced form residuals, i.e. $\widehat{\mathcal{K}}_T = 1/T \sum \text{vech}(\iota_t \iota_t') \otimes \text{vech}(\iota_t \iota_t')$ with $\iota_t = \hat{\Sigma}^{-1/2}u_t$; $K_{n,n}$ is a commutation matrix, which is a $(n^2 \times n^2)$ matrix consisting of $n \times n$ blocks where the $(j,i)-$element of the (i,j) block equals one, elsewhere there are all zeros; and D_n^{+} is the generalized inverse of the duplication matrix D_n.[20] The first bit of the equation (3) represents the sample fourth moments and the second bit the fourth moments implied by a standard normal distribution; τ is the amount of shrinkage that we assign to the normal implied moments; the larger this value the more weight we give to the normality assumption.

It is important to highlight at this point that the sample fourth moments are used only to construct the asymptotic covariance matrix of the reduced form VAR errors volatility matrix

[19]The QML is the maximum estimator of the quasi-likelihood. The quasi-likelihood in this context coincides with the likelihood of the VAR when *incorrectly* assuming normality of the reduced form residuals.

[20]For more details on the notation for the multivariate kurtosis see Kollo (2008).

which measures the asymptotic uncertainty around the consistent estimator of Σ. Fourth sample moments are not used for the shock's identification.

The posterior distribution of the autoregressive parameters conditional on Σ is standard and any prior can be used for the purpose. When there important departures from symmetry, the posterior distributions for the intercept term and the covariance matrix are not independent even for large samples, so robust Bayesian inference requires consistently estimating third moments. As for the fourth moments, we consider the consistent shrinkage estimator given by $\widehat{\mathcal{S}}^{\star}_T = \frac{T}{T+\tau}\widehat{\mathcal{S}}_T$ where $\widehat{\mathcal{S}}_T = (1/T \sum \text{vech}(u_t u_t') \otimes u_t)$.

For inferential purposes it is useful to rewrite the VAR in a seemingly unrelated regression (SUR) format. Let $k = np + 1$, we have

$$\underbrace{Y}_{T\times n} = \underbrace{X}_{T\times k}\underbrace{\Phi}_{k\times n} + \underbrace{E}_{T\times n}$$

$$Y = \begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{pmatrix} = \begin{pmatrix} y_{1,1} & y_{1,2} & \dots & y_{1,n} \\ y_{2,1} & y_{2,2} & \dots & y_{2,n} \\ \vdots & & & \\ y_{T,1} & y_{T,2} & \dots & y_{T,n} \end{pmatrix} \quad X = \begin{pmatrix} \mathbf{x}_0' & 1 \\ \mathbf{x}_1' & 1 \\ \vdots & \\ \mathbf{x}_{T-1}' & 1 \end{pmatrix} \quad \underset{(np\times 1)}{\mathbf{x}_t} = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix}$$

$$\Phi = \begin{pmatrix} \Phi_1' \\ \vdots \\ \Phi_p' \\ \Phi_0' \end{pmatrix} \quad E = \begin{pmatrix} u_1' \\ \vdots \\ u_T' \end{pmatrix}$$

Assuming a flat prior[21], the estimation identification procedure can be then summarized as follows. Let $\widehat{S} = (Y - X\widehat{\Phi})'(Y - X\widehat{\Phi})$ and $\widehat{\Phi} = (X'X)^{-1}X'Y$, the steps of the Gibbs sampler are for $j = 1, ..., J$

1. Draw $\Sigma^{(j)}$ from
$$N\left(\text{vech}(\widehat{S}), \widehat{\mathcal{C}}\right)$$
where $\widehat{\mathcal{C}} = \frac{1}{T}D_n^+\left(\widehat{S}^{1/2} \otimes \widehat{S}^{1/2}\right) D_n \left(\widehat{\mathcal{K}}^{\star} - \text{vech}(I_n)\text{vech}(I_n)'\right) D_n' \left(\widehat{S}^{1/2} \otimes \widehat{S}^{1/2}\right)' D_n^{+'}$ captures the fourth moments.

2. Conditional on $\Sigma^{(j)}$, draw $\Phi^{(j)}$ from
$$N\left(\widehat{\Phi}, \Sigma^{(j)} \otimes (X'X)^{-1}\right)$$
   i. In case of an asymmetric distribution, the intercept, Φ_0, is drawn from
$$N(\widehat{\Phi}_0 + \widehat{\mathcal{S}}^{\star}\widehat{\mathcal{C}}^{-1}\text{vech}(\Sigma^{(j)} - \widehat{S}), \Sigma^{(j)} - 1/T\widehat{\mathcal{S}}^{\star}_T\widehat{\mathcal{C}}^{-1}\widehat{\mathcal{S}}^{\star\prime}_T)$$

3. Draw $\breve{\Omega}$ from a uniform distribution with the Rubio-Ramírez et al. (2010) algorithm and

[21]Se the appendix for extending the Gibb sampler to informative priors.

I. compute the impulse response function and check if the sign restrictions are verified

II. compute the implied structural shocks

$$\breve{\nu}_t^{(j)} = \breve{\Omega}' \left(\Sigma^{(j)}\right)^{-1/2} (y_t - \Phi_1^{(j)} y_{t-1} - ... - \Phi_p^{(j)} y_{t-p} - \Phi_0^{(j)})$$

and check if the higher moment inequality restrictions are satisfied

If both [I] and [II] are satisfied, keep the draw $\Omega^{(j)} = \breve{\Omega}$. Else repeat [I] and [II].

After a suitable number of iterations, the draws are representative of the posterior distribution of interest. The estimation of the reduced form parameters and the computation of the impulse responses and of the higher order moments is performed using the toolbox described in Ferroni and Canova (2021).[22]

A.5.1 Informative priors

Assume a multivariate normal MN prior for the autoregressive parameters

$$\Phi \sim N(\Phi_0, \Sigma \otimes V) \quad =(2\pi)^{-nk/2}|\Sigma|^{-k/2}|V|^{-n/2} \exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}(\Phi - \Phi_0)' V^{-1}(\Phi - \Phi_0)\right]\right\}$$

Modify the second step of the Gibbs sample with the following

$$\Phi|\Sigma, Y, X, \sim N(\overline{\Phi}, \Sigma \otimes (X'X + V^{-1})^{-1})$$
$$\overline{\Phi} = (X'X + V^{-1})^{-1}(X'Y + V^{-1}\Phi_0)$$

[22]Codes for replication can be found on the Github page.