# Bayesian Inference on Structural Impulse Response Functions

Mikkel Plagborg-Møller*

This version: August 26, 2016. First version: October 26, 2015.

**Abstract:** I propose to estimate structural impulse responses from macroeconomic time series by doing Bayesian inference on the Structural Vector Moving Average representation of the data. This approach has two advantages over Structural Vector Autoregressions. First, it imposes prior information directly on the impulse responses in a flexible and transparent manner. Second, it can handle noninvertible impulse response functions, which are often encountered in applications. Rapid simulation of the posterior of the impulse responses is possible using an algorithm that exploits the Whittle likelihood. The impulse responses are partially identified, and I derive the frequentist asymptotics of the Bayesian procedure to show which features of the prior information are updated by the data. The procedure is used to estimate the effects of technological news shocks on the U.S. business cycle.

**Keywords:** Bayesian inference, Hamiltonian Monte Carlo, impulse response function, news shock, nonfundamental, noninvertible, partial identification, structural vector autoregression, structural vector moving average, Whittle likelihood.

## 1 Introduction

Since Sims (1980), Structural Vector Autoregression (SVAR) analysis has been the most popular method for estimating the impulse response functions (IRFs) of observed macro

*Harvard University, email: plagborg@fas.harvard.edu. I am grateful for comments from Isaiah Andrews, Regis Barnichon, Varanya Chaubey, Gabriel Chodorow-Reich, Peter Ganong, Ben Hébert, Christian Matthes, Pepe Montiel Olea, Frank Schorfheide, Elie Tamer, Harald Uhlig, and seminar participants at Cambridge, Chicago Booth, Columbia, Harvard Economics and Kennedy School, MIT, Penn State, Princeton, UCL, UPenn, and the 2015 DAEiNA meeting. I thank Eric Sims for sharing his news shock code and Marco Lippi for providing a key part of the proof of Proposition 3. I am indebted to Gary Chamberlain, Gita Gopinath, Anna Mikusheva, Neil Shephard, and Jim Stock for their help and guidance.

variables to unobserved shocks without imposing a specific equilibrium model structure. However, the SVAR model has two well-known drawbacks. First, the under-identification of the parameters requires researchers to exploit prior information to estimate unknown features of the IRFs. Existing inference methods only exploit certain types of prior information, such as zero or sign restrictions, and these methods tend to implicitly impose unacknowledged restrictions. Second, the SVAR model does not allow for noninvertible IRFs. These can arise when the econometrician does not observe all variables in economic agents' information sets, as in models with news shocks or noisy signals.

I propose a new method for estimating structural IRFs: Bayesian inference on the Structural Vector Moving Average (SVMA) representation of the data. The parameters of this model are the IRFs, so prior information can be imposed by placing a flexible Bayesian prior distribution directly on the parameters of scientific interest. The SVMA approach thus overcomes the two drawbacks of SVAR analysis. First, researchers can flexibly and transparently exploit all types of prior information about IRFs. Second, the SVMA model does not restrict the IRFs to be invertible *a priori*, so the model can be applied to a wider range of empirical questions than the SVAR model. To take the SVMA model to the data, I develop a posterior simulation algorithm that uses the Whittle likelihood approximation to speed up computations. As the IRFs are partially identified, I derive the frequentist asymptotic limit of the posterior distribution to show which features of the prior are dominated by the data.

The first key advantage of the SVMA model is that prior information about IRFs – the parameters of scientific interest – can be imposed in a direct, flexible, and transparent manner. In standard SVAR analysis the mapping between parameters and IRFs is indirect, and the IRFs are estimated by imposing zero or sign restrictions on short- or long-run impulse responses. In the SVMA model the parameters are the IRFs, so all types of prior information about IRFs may be exploited by placing a prior distribution on the parameters. While many prior choices are feasible, I propose a multivariate Gaussian prior that facilitates graphical prior elicitation. In particular, researchers can transparently exploit valuable prior information about the shapes and smoothness of IRFs.

The second key advantage of the SVMA model is that, unlike SVARs, it does not restrict IRFs to be invertible *a priori*, which broadens the applicability of the method. The IRFs are said to be invertible if the current shocks can be recovered as linear functions of current and past – but not future – data. As shown in the literature, *non*invertible IRFs can arise when the econometrician does not observe all variables in the economic agents' information sets, such as in macro models with news shocks or noisy signals. A long-standing problem for standard SVAR methods is that they cannot consistently estimate noninvertible IRFs because the SVAR model implicitly assumes invertibility. Proposed fixes in the SVAR literature either exploit restrictive model assumptions or proxy variables for the shocks, which are not always available. In contrast, the SVMA model is generally applicable since its parametrization does not impose invertibility on the IRFs *a priori*.

To conduct posterior inference about the IRFs, I develop a posterior simulation algorithm that exploits the Whittle (1953) likelihood approximation. Inference in the SVMA model is challenging due to the flexible parametrization, which explains the literature's preoccupation with the computationally convenient SVAR alternative. The computational challenges of the SVMA model are solved by simulating from the posterior using Hamiltonian Monte Carlo (HMC), a Markov Chain Monte Carlo method that is well-suited to high-dimensional models. HMC evaluates the likelihood and score 100,000s of times in realistic applications, so approximating the exact likelihood with the Whittle likelihood drastically reduces computation time. The resulting algorithm is fast, asymptotically efficient, and easy to apply, while allowing for both invertible and noninvertible IRFs. Matlab code available on my website implements all the steps of the inference procedure.[1]

Having established a method for computing the posterior, I derive its frequentist large-sample limit to show how the data updates the prior information. Because the IRFs are partially identified, some aspects of the prior are not dominated by the data in large samples.[2] I establish new results on the frequentist limit of the posterior distribution for a large class of partially identified models under weaker conditions than assumed by Moon & Schorfheide (2012). I then specialize the results to the SVMA model with a non-dogmatic prior, allowing for noninvertible IRFs and non-Gaussian structural shocks. Asymptotically, the role of the data is to pin down the true autocovariances of the data, which in turn pin down the reduced-form (Wold) impulse responses; all other information about structural impulse responses comes from the prior. The result implies that the approximation error incurred by using the Whittle likelihood is negligible asymptotically.

I demonstrate the practical usefulness of the SVMA method in an empirical application that estimates the effects of technological news shocks on the U.S. business cycle. Technological news shocks – signals about future productivity increases – have received much attention in the recent macro literature. My analysis is the first to fully allow for non-

[1] `http://scholar.harvard.edu/plagborg/irf_bayes`

[2] Consistent with Phillips (1989), I use the term "partially identified" in the sense that a nontrivial function of the parameter vector is point identified, but the full parameter vector is not.

invertible IRFs without dogmatically imposing a particular Dynamic Stochastic General Equilibrium (DSGE) model. I use data on productivity, output, and the real interest rate, with the sticky-price DSGE model in E. Sims (2012) serving as a guide to prior elicitation. The results overwhelmingly indicate that the IRFs are noninvertible, implying that no SVAR can consistently estimate the IRFs in this dataset; nevertheless, most IRFs are precisely estimated by the SVMA procedure. The news shock is found to be unimportant for explaining movements in TFP and GDP, but it is an important driver of the real interest rate.

To aid readers who are familiar with SVAR analysis, I demonstrate how to transparently impose standard SVAR identifying restrictions in the SVMA framework, if desired. The SVMA approach can easily accommodate exclusion and sign restrictions on short- and long-run (i.e., cumulative) impulse responses. Prior information can be imposed dogmatically (i.e., with 100% certainty) or non-dogmatically. External instruments can be exploited in the SVMA framework, as in SVARs.

The SVMA estimation approach in this paper is more flexible than previous attempts in the literature, and it appears to be the first method for conducting valid inference about possibly noninvertible IRFs. Hansen & Sargent (1981) and Ito & Quah (1989) estimate SVMA models without assuming invertibility by maximizing the Whittle likelihood, but the only prior information they consider is a class of exact restrictions implied by rational expectations. Barnichon & Matthes (2016) propose a Bayesian approach to inference in SVMA models, but they restrict attention to recursively identified models and they center the prior at SVAR-implied IRFs. None of these three papers develop valid procedures for doing inference on IRFs that may be partially identified and noninvertible.[3] Moreover, each of the three papers imposes parametric functional forms on the IRFs, while I show how to maintain computational tractability with a potentially unrestricted parameter space.

A few SVAR papers have attempted to exploit general types of prior information about IRFs, but these methods are less flexible than the SVMA approach. Furthermore, by assuming an underlying SVAR model, they automatically rule out noninvertible IRFs. Dwyer (1998) works with an inflexible trinomial prior on IRFs. Gordon & Boccanfuso (2001) translate a prior on IRFs into a "best-fitting" prior on SVAR parameters, but Kocięcki (2010) shows that their method neglects the Jacobian of the transformation. Kocięcki's fix requires the transformation to be one-to-one, which limits the ability to exploit prior information

[3] Standard errors in Hansen & Sargent (1981) are only valid when the prior restrictions point identify the IRFs. Barnichon & Matthes (2016) approximate the SVMA likelihood using an autoregressive formula that is explosive when the IRFs are noninvertible, causing serious numerical instability. Barnichon & Matthes focus on invertible IRFs and extend the model to allow for asymmetric and state-dependent effects of shocks.

about long-run responses, shapes, and smoothness. Baumeister & Hamilton (2015b), who improve on the method of Sims & Zha (1998), persuasively argue for an explicit Bayesian approach to imposing prior information. Their Bayesian SVAR method allows for a fully flexible prior on impact impulse responses, but they assume invertibility, and their prior on longer-horizon impulse responses is implicit and chosen for computational convenience.

Section 2 reviews SVARs and then discusses the SVMA model, invertibility, identification, and prior elicitation. Section 3 outlines the posterior simulation method. Section 4 empirically estimates the role of technological news shocks in the U.S. business cycle. Section 5 contains asymptotic analysis. Section 6 shows that popular SVAR restrictions can be imposed in the SVMA framework. Section 7 suggests topics for future research. Applied readers may want to focus on Sections 2 to 4. Technical details are relegated to Appendix A; in particular, notation is defined in Appendix A.1. Proofs can be found in Appendix B. The Online Appendix on the author's website includes an illustrative simulation study, supplemental empirical results, and further technical details (see Footnote 1).

# 2 Model, invertibility, and prior elicitation

In this section I describe the SVMA model and my method for imposing priors on IRFs. I define the SVMA model, whose parameters are IRFs. Because the SVMA model does not restrict the IRFs to be invertible, it can be applied to more empirical settings than the SVAR approach. The lack of identification of the IRFs necessitates the use of prior information, which I impose by placing a transparent and flexible prior distribution on the IRFs.

## 2.1 SVARs and their shortcomings

I begin with a brief review of Structural Vector Autoregressions (SVARs). The parametrization of the SVAR model makes it difficult to transparently exploit certain types of valuable prior information about impulse responses. Moreover, SVARs are ill-suited for empirical applications in which the econometrician has less information than economic agents.

Modern dynamic macroeconomics attaches primary importance to *impulse response functions* (IRFs). The economy is assumed to be driven by unpredictable shocks (impulses) whose effect on observable macro aggregates is known as the propagation mechanism. Hansen & Sargent (1981) and Watson (1994, Sec. 4) argue that – in a linear setting – this impulse-propagation paradigm is captured by the Structural Vector Moving Average (SVMA) model

$$y_t = \Theta(L)\varepsilon_t, \quad \Theta(L) = \sum_{\ell=0}^{\infty} \Theta_\ell L^\ell, \tag{1}$$

where $L$ denotes the lag operator, $y_t = (y_{1,t}, \dots, y_{n,t})'$ is an $n$-dimensional vector of observed macro variables, and the structural shocks $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})'$ form a martingale difference sequence with $E(\varepsilon_t \varepsilon_t') = \text{diag}(\sigma)^2$, $\sigma = (\sigma_1, \dots, \sigma_n)'$. Most linearized discrete-time macro models can be written in SVMA form. $\Theta_{ij,\ell}$, the $(i, j)$ element of $\Theta_\ell$, is the *impulse response* of variable $i$ to shock $j$ at horizon $\ell$ after the shock's initial impact. The IRF $(\Theta_{ij,\ell})_{\ell \geq 0}$ is thus a key object of scientific interest in macroeconomics (Ramey, 2016).

For computational reasons, most researchers follow Sims (1980) and estimate structural IRFs using a SVAR model

$$A(L)y_t = H\varepsilon_t, \quad A(L) = I_n - \sum_{\ell=1}^{m} A_\ell L^\ell, \tag{2}$$

where $m$ is a finite lag length, and the matrices $A_1, \dots, A_m$ and $H$ are each $n \times n$. The SVAR and SVMA models are closely related: If the SVAR is stable – i.e., the polynomial $A(L)$ has a one-sided inverse – the SVAR model (2) implies that the data has an SVMA representation (1) with IRFs given by $\Theta(L) = \sum_{\ell=0}^{\infty} \Theta_\ell L^\ell = A(L)^{-1}H$. The SVAR model is computationally attractive because the parameters $A_\ell$ are regression coefficients.

The IRFs implied by the SVAR model are not identified from the data if the shocks are unobserved, as is usually the case.[4] While the VAR polynomial $A(L)$ can be recovered from a regression of $y_t$ on its lags, the impact matrix $H$ and shock standard deviations $\sigma$ are not identified.[5] As knowledge of $H$ is required to pin down the SVAR IRFs, the latter are under-identified. Thus, the goal of the SVAR literature is to exploit weak prior information about the model parameters to estimate unknown features of the IRFs.

One drawback of the SVAR model is that its parametrization makes it difficult to transparently exploit certain types of prior information. The IRFs $\Theta(L) = A(L)^{-1}H$ implied by the SVAR model are nonlinear functions of the parameters $(A(L), H)$, and impulse responses

[4] If the structural shocks $\varepsilon_t$ were known, the IRFs in the SVMA model (1) could be estimated by direct regressions of $y_t$ on lags of $\varepsilon_t$ (Jordà, 2005).

[5] Denote the reduced-form (Wold) forecast error by $u_{t|t-1} = y_t - \text{proj}(y_t \mid y_{t-1}, y_{t-2}, \dots) = H\varepsilon_t$, where "proj" denotes population linear projection. Let $E(u_{t|t-1}u_{t|t-1}') = JJ'$ be the (identified) Cholesky decomposition of the forecast error covariance matrix. Then all that the second moments of the data reveal about $H$ and $\sigma$ is that $H\,\text{diag}(\sigma) = JQ$ for some unknown $n \times n$ orthogonal matrix $Q$ (Uhlig, 2005, Prop. A.1).

$\Theta_{ij,\ell}$ at long horizons $\ell$ are extrapolated from short-run parameters. Hence, the shapes and smoothness of the model-implied IRFs depend indirectly on the SVAR parameters, which impedes the use of prior information about such features of the IRFs.[6] Instead, SVAR papers impose zero or sign restrictions on short- or long-run impulse responses to sharpen identification.[7] Because of the indirect parametrization, such SVAR identification schemes are known to impose additional unintended and unacknowledged prior information about IRFs.[8]

A second drawback of the SVAR model is the invertibility problem. The defining property of the SVAR model (2) is that the structural shocks $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})'$ can be recovered linearly from the history $(y_t, y_{t-1}, \dots)$ of observed data, given knowledge of $H$ and $\sigma$. This *invertibility* assumption – that future data is not required to recover the current shocks – is arbitrary and may be violated if the econometrician does not observe all variables relevant to the decisions of forward-looking economic agents. Section 2.3 discusses invertibility in greater detail and provides references.

I overcome the drawbacks of the SVAR model by doing Bayesian inference directly on the SVMA model (1). Since the parameters of this model are the IRFs themselves, prior information can be imposed directly on the objects of scientific interest. Moreover, there is no need to restrict the IRFs to the invertible part of the parameter space.

## 2.2 SVMA model

I now discuss the SVMA model in detail and show that its parameters can be interpreted as IRFs. Then I illustrate the natural parametrization by example.

The SVMA model assumes the observed time series $y_t = (y_{1,t}, \dots, y_{n,t})'$ are driven by current and lagged values of unobserved, unpredictable shocks $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})'$ (Hansen & Sargent, 1981). For simplicity, I follow the SVAR literature in assuming that the number $n$ of shocks is known and equals the number of observed series. However, most methods in this paper generalize to the case with more shocks than variables, see Section 7.

[6]The shapes of the IRFs are governed by the magnitudes and imaginary parts of the roots of the VAR lag polynomial $A(L)$, and the roots are in turn complicated functions of the lag matrices $A_1, \dots, A_m$.

[7]Ramey (2016) and Stock & Watson (2016) review SVAR identification schemes.

[8]Consider the AR(1) model $y_t = A_1 y_{t-1} + \varepsilon_t$ with $n = m = 1$ and $|A_1| < 1$. The IRF is $\Theta_\ell = A_1^\ell$, so the sign restriction $\Theta_1 \geq 0$ implicitly also restricts $\Theta_\ell \geq 0$ for all $\ell \geq 2$. Increasing the lag length $m$ makes the model more flexible but the mapping from parameters to IRFs more complicated. Baumeister & Hamilton (2015b) and Arias, Rubio-Ramírez & Waggoner (2016) highlight other issues for SVAR inference procedures.

**Assumption 1** (SVMA model)**.**

$$y_t = \Theta(L)\varepsilon_t,\ t \in \mathbb{Z}, \quad \Theta(L) = \sum_{\ell=0}^{q} \Theta_\ell L^\ell, \tag{3}$$

*where $L$ is the lag operator, $q$ is the finite MA lag length, and $\Theta_0, \Theta_1, \dots, \Theta_q$ are each $n \times n$ coefficient matrices. The shocks are serially and mutually unpredictable: For each $t$ and $j$, $E(\varepsilon_{j,t} \mid \{\varepsilon_{k,t}\}_{k \neq j}, \{\varepsilon_s\}_{-\infty<s<t}) = 0$ and $E(\varepsilon_{j,t}^2) = \sigma_j^2$, where $\sigma_j > 0$.*

For simplicity, I assume that the moving average (MA) lag length $q$ is finite and known, but it is of course possible to estimate $q$ using information criteria, cf. Section 4. To fit persistent data $q$ must be large, which the computational strategy in Section 3 is well-suited for. The methods in this paper can be extended to the case $q = \infty$, as briefly discussed in Section 7. The assumption that $y_t$ has mean zero is made for notational convenience.

The SVMA and SVAR models are related but not equivalent. If the matrix lag polynomial $\Theta(L)$ has a one-sided inverse $D(L) = \sum_{\ell=0}^{\infty} D_\ell L^\ell = \Theta(L)^{-1}$, the SVMA model (3) is compatible with an underlying SVAR $D(L)y_t = \varepsilon_t$ (with lag length $m = \infty$). However, the fact that I do not constrain $\Theta(L)$ to have a one-sided inverse is key to allowing for noninvertible IRFs, cf. Section 2.3. Assumption 1 imposes stationary, linear dynamics, which is restrictive but standard in the SVAR literature. The condition that $\varepsilon_t$ is a martingale difference sequence with mutually unpredictable components is also standard and operationalizes the interpretation of $\varepsilon_t$ as a vector of conceptually independent structural shocks.

Unlike in SVARs, the parameters of the SVMA model have direct economic interpretations as impulse responses (see also Barnichon & Matthes, 2016). Denote the $(i, j)$ element of matrix $\Theta_\ell$ by $\Theta_{ij,\ell}$. The index $\ell$ will be referred to as the horizon. For each $j \in \{1, \dots, n\}$, choose an $i_j \in \{1, \dots, n\}$ and normalize the impact response of variable $i_j$ to shock $j$: $\Theta_{i_j j,0} = 1$. Then the parameter $\Theta_{ij,\ell}$ is the expected response at horizon $\ell$ of variable $i$ to shock $j$, for a shock magnitude that raises variable $i_j$ by one unit on impact:[9]

$$\Theta_{ij,\ell} = E(y_{i,t+\ell} \mid \varepsilon_{j,t} = 1) - E(y_{i,t+\ell} \mid \varepsilon_{j,t} = 0). \tag{4}$$

The *impulse response function* (IRF) of variable $i$ to shock $j$ is the $(q+1)$-dimensional vector $(\Theta_{ij,0}, \Theta_{ij,1}, \dots, \Theta_{ij,q})'$. In addition to the impulse response parameters $\Theta_{ij,\ell}$, the model contains the shock standard deviation parameters $\sigma_j$, which govern the overall magnitudes of the responses to one-standard-deviation impulses to $\varepsilon_{j,t}$.

[9]Henceforth, moments of the data and shocks are implicitly conditioned on the parameters $(\Theta, \sigma)$.

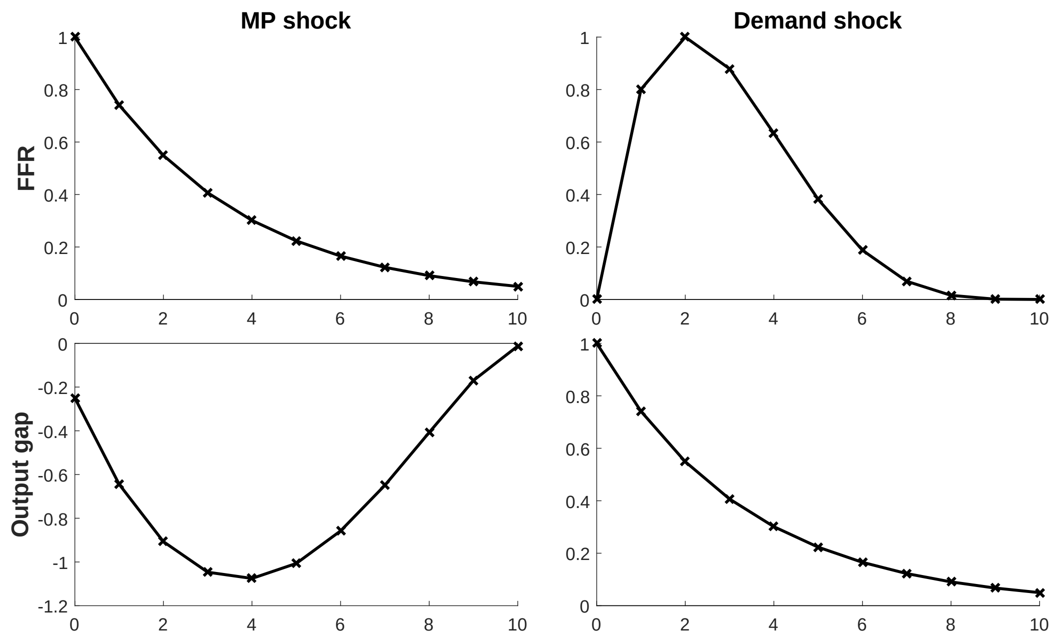


**Figure 1:** Hypothetical IRFs of two observed variables (along rows) to two unobserved shocks (along columns). The upper right display, say, shows the IRF of the FFR to the demand shock. The horizontal axes represent the impulse response horizon $\ell = 0, 1, \ldots, q$, where $q = 10$. IRFs in the left column are normalized so a positive monetary policy (MP) shock yields a 100 basis point increase in the FFR on impact; IRFs in the right column are normalized so a positive demand shock yields a 1 percentage point increase in the output gap on impact.

The parameters are best understood through an example. Figure 1 plots a hypothetical set of impulse responses for a bivariate application with two observed time series, the federal funds rate (FFR) $y_{1,t}$ and the output gap $y_{2,t}$, and two unobserved shocks, a monetary policy shock $\varepsilon_{1,t}$ and a demand shock $\varepsilon_{2,t}$. I impose the normalizations $i_1 = 1$ and $i_2 = 2$, so that $\Theta_{21,3}$, say, is the horizon-3 impulse response of the output gap to a monetary policy shock that raises the FFR by 1 unit (100 basis points) on impact. Each impulse response (the crosses in the figure) corresponds to a distinct IRF parameter $\Theta_{ij,\ell}$. The joint visualization of these parameters is familiar from theoretical macro modeling. In contrast, parameters in SVAR models are indirectly related to IRFs and do not carry graphical intuition in and of themselves. The natural and flexible parametrization of the SVMA model facilitates the incorporation of prior information about IRFs, as described below.

Because I wish to estimate the IRFs using parametric Bayesian methods, it is necessary to strengthen Assumption 1 by assuming a specific distribution for the shocks $\varepsilon_t$. For concreteness I impose the working assumption that they are i.i.d. Gaussian.

**Assumption 2** (Gaussian shocks)**.** $\varepsilon_t \overset{i.i.d.}{\sim} N(0, \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)),\ t \in \mathbb{Z}$.

The Gaussianity assumption places the focus on the unconditional second-order properties of the data $y_t$, as is standard in the SVAR literature, but the assumption is not central to my analysis. Section 5 shows that if the Bayesian posterior distribution for the IRFs is computed under Assumption 2 and a non-dogmatic prior distribution, the large-sample limit of the posterior is the same whether or not the data distribution is truly Gaussian. Moreover, the method for sampling from the posterior in Section 3 is flexible and can be adapted to non-Gaussian and/or heteroskedastic likelihoods.

## 2.3 Invertibility

One advantage of the SVMA model is that it allows for noninvertible IRFs. These arise frequently in applications in which the econometrician does not observe all variables in economic agents' information sets. Because the SVMA model does not restrict IRFs to be invertible *a priori*, it is more broadly applicable than the SVAR model.

The IRF parameters are *invertible* if the current shock $\varepsilon_t$ can be recovered as a linear function of current and past – but not future – values $(y_t, y_{t-1}, \dots)$ of the observed data, given knowledge of the parameters.[10] In this sense, *non*invertibility is caused by economically important variables being omitted from the econometrician's information set.[11] Invertibility is a property of the collection of $n^2$ IRFs, and an invertible collection of IRFs can be rendered noninvertible by removing or adding observed variables or shocks.

Invertibility is not a compelling *a priori* restriction when estimating structural IRFs, for two reasons. First, the definition of invertibility is statistically motivated and has little economic content. For example, the reasonable-looking IRFs in Figure 1 happen to be noninvertible, but minor changes to the lower left IRF in the figure render the IRFs invertible. Second, interesting macro models generate noninvertible IRFs, such as models with news shocks or noisy signals.[12] Intuitively, upon receiving a signal about changes in policy or economic fundamentals that will occur sufficiently far into the future, economic agents change their current behavior much less than their future behavior. Thus, future – in addition to

[10] Precisely, the IRFs are invertible if $\varepsilon_t$ lies in the closed linear span of $(y_t, y_{t-1}, \dots)$. Invertible MA representations are also referred to as "fundamental" in the literature. See Hansen & Sargent (1981, 1991) and Lippi & Reichlin (1994) for extensive mathematical discussions of invertibility in SVMAs and SVARs.

[11] See Hansen & Sargent (1991), Sims & Zha (2006), Fernández-Villaverde, Rubio-Ramírez, Sargent & Watson (2007), Forni, Giannone, Lippi & Reichlin (2009), Leeper, Walker & Yang (2013), Forni, Gambetti & Sala (2014), and Lütkepohl (2014).

[12] See Alessi, Barigozzi & Capasso (2011, Sec. 4–6), Blanchard, L'Huillier & Lorenzoni (2013, Sec. II), Leeper et al. (2013, Sec. 2), and Beaudry & Portier (2014, Sec. 3.2).

current and past – data is needed to distinguish the signal from other concurrent shocks.

By their very definition, SVARs implicitly restrict IRFs to be invertible, as discussed in Section 2.1. No SVAR identification strategy can therefore consistently estimate noninvertible IRFs. This fact has spawned an extensive literature trying to salvage the SVAR approach. Some papers assume additional model structure,[13] while others rely on the availability of proxy variables for the shocks.[14] These methods only produce reliable results under additional assumptions or if the requisite data is available, whereas the SVMA approach always yields correct inference about IRFs regardless of invertibility. If available, proxy variables can be incorporated in SVMA analysis to improve identification.

The SVMA model (3) is parametrized directly in terms of IRFs and does not impose invertibility *a priori* (Hansen & Sargent, 1981). Specifically, the IRFs are invertible if and only if the polynomial $z \mapsto \det(\Theta(z))$ has no roots inside the unit circle.[15] In general, the structural shocks can be recovered from past, current, and *future* values of the data:[16]

$$\varepsilon_t = D(L)y_t, \quad D(L) = \sum_{\ell=-\infty}^{\infty} D_\ell L^\ell = \Theta(L)^{-1}.$$

Under Assumption 1, the structural shocks can thus be recovered from *multi-step* forecast errors: $\varepsilon_t = \sum_{\ell=0}^{\infty} D_\ell u_{t+\ell|t-1}$, where $u_{t+\ell|t-1} = y_{t+\ell} - \text{proj}(y_{t+\ell} \mid y_{t-1}, y_{t-2}, \dots)$ is the econometrician's $(\ell+1)$-step error. Only if the IRFs are invertible do we have $D_\ell = 0$ for $\ell \geq 1$, in which case $\varepsilon_t$ is a linear function of the *one-step* (Wold) error $u_{t|t-1}$, as SVARs assume.

As an illustration, consider a univariate SVMA model with $n = q = 1$:

$$y_t = \varepsilon_t + \Theta_1 \varepsilon_{t-1}, \quad \Theta_1 \in \mathbb{R}, \quad E(\varepsilon_t^2) = \sigma^2. \tag{5}$$

If $|\Theta_1| \leq 1$, the IRF $\Theta = (1, \Theta_1)$ is invertible: The shock has the SVAR representation

[13] Lippi & Reichlin (1994) and Klaeffing (2003) characterize the range of noninvertible IRFs consistent with a given estimated SVAR, while Mertens & Ravn (2010) and Forni, Gambetti, Lippi & Sala (2013) select a single such IRF using additional model restrictions. Lanne & Saikkonen (2013) develop asymptotic theory for a modified VAR model that allows for noninvertibility, but they do not consider structural estimation.

[14] Sims & Zha (2006), Fève & Jidoud (2012), Sims (2012), Beaudry & Portier (2014, Sec. 3.2), and Beaudry, Fève, Guay & Portier (2015) argue that noninvertibility need not cause large biases in SVAR estimation if forward-looking variables are available. Forni et al. (2009) and Forni et al. (2014) use information from large panel data sets to ameliorate the omitted variables problem; based on the same idea, Giannone & Reichlin (2006) and Forni & Gambetti (2014) propose tests of invertibility.

[15] That is, if and only if $\Theta(L)^{-1}$ is a one-sided lag polynomial, so that the SVAR representation $\Theta(L)^{-1}y_t = \varepsilon_t$ obtains (Brockwell & Davis, 1991, Thm. 11.3.2, and Remark 1, p. 128).

[16] See Brockwell & Davis (1991, Thm. 3.1.3) and Lippi & Reichlin (1994, p. 312). The matrix lag polynomial $D(L) = \Theta(L)^{-1}$ is not well-defined in the knife-edge case $\det(\Theta(1)) = \det(\sum_{\ell=0}^{q} \Theta_\ell) = 0$.

$\varepsilon_t = \sum_{\ell=0}^{\infty}(-\Theta_1)^\ell y_{t-\ell}$, so it can be recovered using current and past values of the data. In contrast, if $|\Theta_1| > 1$, no SVAR representation for $\varepsilon_t$ exists: $\varepsilon_t = -\sum_{\ell=1}^{\infty}(-\Theta_1)^{-\ell} y_{t+\ell}$, so *future* values of the data are required to recover the current structural shock. The latter case is consistent with the SVMA model (5) but inconsistent with any SVAR model (2).[17]

Bayesian analysis of the SVMA model can be carried out without reference to the invertibility of the IRFs. The formula for the Gaussian SVMA likelihood function is the same in either case, and standard state-space methods can be used to estimate the structural shocks, cf. Sections 3 and 4 and Hansen & Sargent (1981). This contrasts sharply with SVAR analysis, where special tools are needed to handle noninvertible specifications. Since invertibility is a rather arcane issue without much economic content, it is helpful that the SVMA model allows the researcher to focus on matters that do have economic significance.

## 2.4 Identification

The IRFs in the SVMA model are only partially identified, as in SVAR analysis. The lack of identification arises because the model treats all shocks symmetrically and because noninvertible IRFs are not ruled out *a priori*.

Any two sets of IRFs that give rise to the same autocovariance function (ACF) are observationally equivalent, assuming Gaussian shocks. Under Assumption 1, the matrix ACF of the time series $\{y_t\}$ is given by

$$\Gamma(k) = E(y_{t+k}y_t') = \begin{cases} \sum_{\ell=0}^{q-k} \Theta_{\ell+k} \operatorname{diag}(\sigma)^2 \Theta_\ell' & \text{if } 0 \le k \le q, \\ 0 & \text{if } k > q. \end{cases} \tag{6}$$

Under Assumptions 1 and 2, the ACF completely determines the distribution of the observed mean-zero strictly stationary Gaussian time series $y_t$. The identified set $\mathcal{S}$ for the IRF parameters $\Theta = (\Theta_0, \Theta_1, \dots, \Theta_q)$ and shock standard deviation parameters $\sigma = (\sigma_1, \dots, \sigma_n)'$ is then a function of the ACF:

$$\mathcal{S}(\Gamma) = \left\{ (\tilde{\Theta}_0, \dots, \tilde{\Theta}_q) \in \Xi_\Theta, \tilde{\sigma} \in \Xi_\sigma \colon \sum_{\ell=0}^{q-k} \tilde{\Theta}_{\ell+k} \operatorname{diag}(\tilde{\sigma})^2 \tilde{\Theta}_\ell' = \Gamma(k), 0 \le k \le q \right\},$$

where $\Xi_\Theta = \{(\tilde{\Theta}_0, \dots, \tilde{\Theta}_q) \in \mathbb{R}^{n \times n(q+1)} \colon \tilde{\Theta}_{i_j j,0} = 1, 1 \le j \le n\}$ is the parameter space for

[17] If $|\Theta_1| > 1$, an SVAR (with $m = \infty$) applied to the time series (5) estimates the incorrect invertible IRF $(1, 1/\Theta_1)$ and (Wold) "shock" $u_{t|t-1} = \varepsilon_t + (1 - \Theta_1^2)\sum_{\ell=1}^{\infty}(-\Theta_1)^{-\ell}\varepsilon_{t-\ell}$.

$\Theta$, and $\Xi_\sigma = \{(\tilde{\sigma}_1, \dots, \tilde{\sigma}_n)' \in \mathbb{R}^n \colon \tilde{\sigma}_j > 0, 1 \le j \le n\}$ is the parameter space for $\sigma$.[18] By definition, two parameter configurations contained in the same identified set give rise to the same value of the SVMA likelihood function under Gaussian shocks.

The identified set for the SVMA parameters is large in economic terms. Appendix A.2 provides a constructive characterization of $\mathcal{S}(\Gamma)$, building on Hansen & Sargent (1981) and Lippi & Reichlin (1994). I summarize the main insights here.[19] The identified set contains uncountably many parameter configurations if the number $n$ of shocks exceeds 1. The lack of identification is not just a technical curiosity but is of primary importance to economic conclusions. For example, as in SVARs, for any observed ACF $\Gamma(\cdot)$, any horizon $\ell$, any shock $j$, and any variable $i \neq i_j$, there exist IRFs in the identified set $\mathcal{S}(\Gamma)$ with $\Theta_{ij,\ell} = 0$.

One reason for under-identification, also present in SVARs (cf. Section 2.1), is that the assumptions so far treat the $n$ shocks symmetrically: Without further restrictions, the model and data do not distinguish the first shock from the second shock, say. Precisely, the two parameter configurations $(\Theta, \sigma)$ and $(\tilde{\Theta}, \tilde{\sigma})$ lie in the same identified set if there exists an orthogonal $n \times n$ matrix $Q$ such that $\tilde{\Theta} \operatorname{diag}(\tilde{\sigma}) Q = \Theta \operatorname{diag}(\sigma)$.

The second source of under-identification is that the SVMA model, unlike SVARs, does not arbitrarily restrict the IRFs to be invertible. For any noninvertible set of IRFs there always exists an observationally equivalent invertible set of IRFs (if $n > 1$, there exist several). If $nq > 1$, there are also several other observationally equivalent noninvertible IRFs. This identification issue arises even if, say, we impose exclusion restrictions on the elements of $\Theta_0$ to exactly identify the orthogonal matrix $Q$ in the previous paragraph.

Figure 2 illustrates the identification problem due to noninvertibility for a univariate model with $n = 1$ and $q = 4$: $y_t = \varepsilon_t + \sum_{\ell=1}^{4} \Theta_\ell \varepsilon_{t-\ell}$, $\Theta_\ell \in \mathbb{R}$, $E(\varepsilon_t^2) = \sigma^2$. The ACF in the left panel of the figure is consistent with the four IRFs shown in the right panel. The invertible IRF (thick line) is the one that would be estimated by a SVAR (with lag length $m = \infty$). Yet there exist three other IRFs that have very different economic implications but are equally consistent with the observed ACF.[20] If $n > 1$, the identification problem is even

[18] If the shocks $\varepsilon_t$ were known to have a non-Gaussian distribution, the identified set would change due to the additional information provided by higher-order moments of the data, cf. Section 5.3.

[19] The identification problem is not easily cast in the framework of interval identification, as $\mathcal{S}(\Gamma)$ is of strictly lower dimension than the parameter space $\Xi_\Theta \times \Xi_\sigma$. Still, expression (6) for $\operatorname{diag}(\Gamma(0))$ implies that the identified set for scaled impulse responses $\Psi_{ij,\ell} = \Theta_{ij,\ell}\sigma_j$ is bounded.

[20] Similarly, in the case $n = q = 1$, the parameters $(\Theta_1, \sigma)$ yield the same ACF as the parameters $(\tilde{\Theta}_1, \tilde{\sigma})$, where $\tilde{\Theta}_1 = 1/\Theta_1$ and $\tilde{\sigma} = \sigma\Theta_1$. If $|\Theta_1| \le 1$, an SVAR would estimate the invertible IRF $(1, \Theta_1)$ for which most of the variation in $y_t$ is due to the current shock $\varepsilon_t$. But the data would be equally consistent with the noninvertible IRF $(1, \tilde{\Theta}_1)$ for which $y_t$ is mostly driven by the previous shock $\varepsilon_{t-1}$.

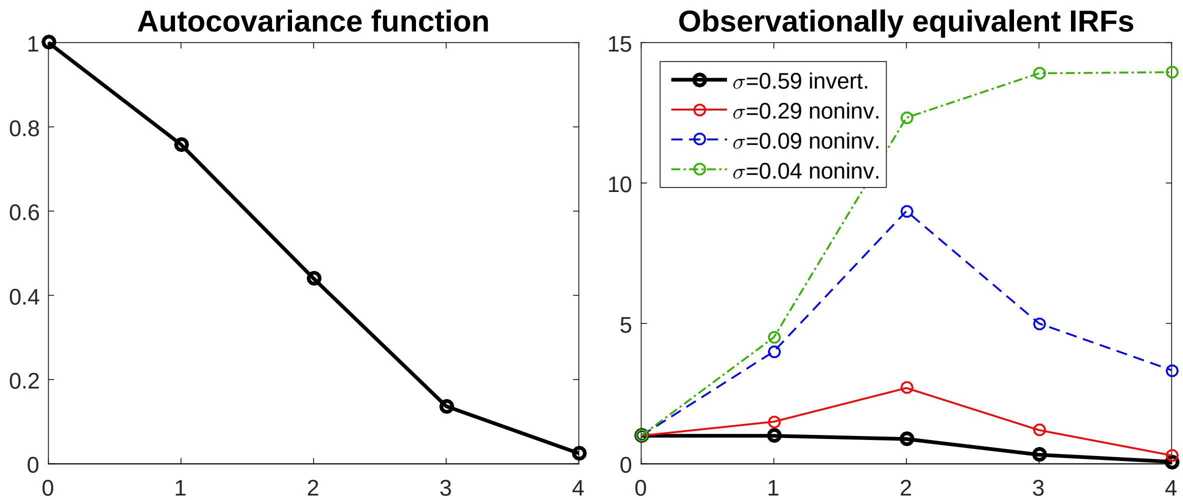


**Figure 2:** Example of IRFs that generate the same ACF, based on a univariate SVMA model with $n = 1$ and $q = 4$. The right panel shows the four IRFs that generate the particular ACF in the left panel; associated shock standard deviations are shown in the figure legend.

more severe, as described in Appendix A.2. Hence, to learn anything useful about unknown features of the IRFs, researchers must exploit available prior information.

## 2.5 Prior specification and elicitation

In addition to handling noninvertible IRFs, the other key advantage of the SVMA model is its natural parametrization, which allows prior information to be imposed directly on the IRFs. Prior information sharpens identification and disciplines the flexible IRF parametrization. Researchers often have access to more prior information about IRFs than what SVAR methods exploit. I propose a transparent procedure for imposing all types of prior information about IRFs in a unified way.

TYPES AND SOURCES OF PRIOR INFORMATION. To impose prior information, the researcher must have some knowledge about the identity and effects of the unobserved shocks. As in SVAR analysis, the researcher postulates that, say, the first shock $\varepsilon_{1,t}$ is a monetary policy shock, the second shock $\varepsilon_{2,t}$ is a demand shock, etc. Then prior information about the effects of the shocks, i.e., the IRFs, is imposed. Prior information can be dogmatic (100% certain, as is common in SVAR analysis) or non-dogmatic (less than 100% certain).

Because the SVMA model is parametrized in terms of IRFs, it is possible to exploit many types of prior information. Researchers often have some prior information about *magnitudes* of certain impulse responses. For example, the impact response of the output gap to a monetary policy shock that lowers the FFR by 100 basis points is unlikely to exceed 2 percentage points. Researchers typically have more informative priors about the *signs* of certain impulse responses. Researchers may also have beliefs about the *shapes* of IRFs, e.g., whether they are likely to be monotonic or hump-shaped (i.e., the effect gradually builds up and then peters out). Finally, researchers often have strong beliefs about the *smoothness* of IRFs, due to adjustment costs, implementation lags, and information frictions.

Prior information may arise from several sources, all of which can be integrated in the graphical prior elicitation procedure introduced below. First, researchers may be guided by structural macroeconomic models whose deep parameters have been calibrated to micro data. Parameter and model uncertainty forbid treating model-implied IRFs as truth, but these may nevertheless be judged to be *a priori* likely, as in the empirical application in Section 4. Second, economic intuition and stylized models yield insight into the likely signs, shapes, and smoothness of the IRFs. Third, microeconometric evidence or macroeconometric studies on related datasets may provide relevant information.

Bayesian approach. Bayesian inference is a unified way to exploit all types of prior information about the IRFs $\Theta$. I place an informative, flexible prior distribution on the SVMA model parameters, i.e., the IRFs $\Theta$ and shock standard deviations $\sigma$. Since there is no known flexible conjugate prior for MA models, I use simulation methods for conducting posterior inference about the structural parameters, as described in Section 3.

The first role of the prior is to attach weights to parameter values that are observationally equivalent based on the data but distinguishable based on prior information, as sketched in Figure 3. The information in the prior and the data is synthesized in the posterior density, which is proportional to the product of the prior density and the likelihood function. As discussed in Section 2.4, the likelihood function does not have a unique maximum due to partial identification. The SVMA analysis thus depends crucially on the prior information imposed, just as SVAR analysis depends on the identification scheme. The frequentist asymptotics in Section 5 show formally that only some features of the prior information can be updated and falsified by the data. This is unavoidable due to the lack of identification (Poirier, 1998), but it does underscore the need for a transparent and flexible prior elicitation procedure.

The second role of the prior is to discipline the flexible IRF parametrization. IRFs are high-dimensional objects, so prior information about their magnitudes, shapes, or smoothness is necessary to avoid overfitting in realistic data samples. In comparison, SVARs achieve

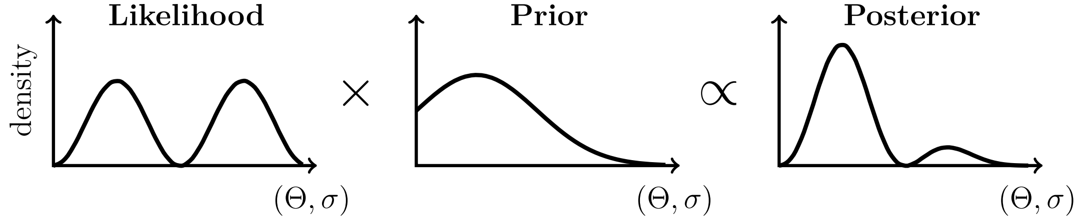


**Figure 3:** Conceptual illustration of how the likelihood function and the prior density combine to yield the posterior density. Even though the likelihood has multiple peaks of equal height, the posterior may be almost unimodal, depending on the strength of prior information.

dimension reduction by rigidly parametrizing the IRFs, effectively extrapolating long-run responses from short-run correlations in the data.

Gaussian prior. While many priors are possible, I first discuss an especially transparent multivariate Gaussian prior distribution. Under Gaussianity, the prior hyperparameters are easily visualized, as illustrated by example below. However, I stress that neither the overall SVMA approach nor the numerical methods in this paper rely on Gaussianity of the prior. I describe other possible prior choices below.

The multivariate Gaussian prior distribution on the impulse responses is given by

$$\Theta_{ij,\ell} \sim N(\mu_{ij,\ell}, \tau_{ij,\ell}^2), \quad 0 \leq \ell \leq q,$$

$$\text{Corr}(\Theta_{ij,\ell+k}, \Theta_{ij,\ell}) = \rho_{ij}^k, \quad 0 \leq \ell \leq \ell + k \leq q, \tag{7}$$

for each $(i, j)$. This correlation structure means that the prior smoothness of IRF $(i, j)$ is governed by $\rho_{ij}$, as illustrated below. For simplicity, the IRFs $(\Theta_{ij,0}, \Theta_{ij,1}, \ldots, \Theta_{ij,q})$ are *a priori* independent across $(i, j)$ pairs. The normalized impulse responses have $\mu_{i_j j,0} = 1$ and $\tau_{i_j j,0} = 0$ for each $j$. The shock standard deviations $\sigma_1, \ldots, \sigma_n$ are *a priori* mutually independent and independent of the IRFs, with prior marginal distribution

$$\log \sigma_j \sim N(\mu_j^\sigma, (\tau_j^\sigma)^2)$$

for each $j$. In practice, the prior variances $(\tau_j^\sigma)^2$ for the log shock standard deviations can be chosen to be a large number. Because the elements of $\sigma$ scale the ACF, which is identified, the data will typically be quite informative about the standard deviations of the shocks, provided that the prior on the IRFs is sufficiently informative.

The key hyperparameters in this Gaussian prior are the prior means $\mu_{ij,\ell}$ and variances

$\tau^2_{ij,\ell}$ of each impulse response, and the prior smoothness hyperparameter $\rho_{ij}$ for each IRF. The prior means and variances can be elicited graphically by drawing a figure with a "best guess" for each IRF and then placing a 90% (say) prior confidence band around each IRF. Once these hyperparameters have been elicited, the prior smoothness $\rho_{ij}$ of each IRF can be elicited by trial-and-error simulation from the multivariate Gaussian prior.

The prior elicitation process is illustrated in Figures 4 and 5, which continue the bivariate example from Figure 1. The figures show a choice of prior means and 90% prior confidence bands for each of the impulse responses, directly implying corresponding values for the $\mu_{ij,\ell}$ and $\tau^2_{ij,\ell}$ hyperparameters. The prior distributions in the figures embed many different kinds of prior information. For example, the IRF of the FFR to a positive demand shock is believed to be hump-shaped with high probability, and the IRF of the output gap to a contractionary monetary policy shock is believed to be negative at horizons 2–8 with high probability. Yet the prior expresses substantial uncertainty about several of the impulse responses.

After having elicited the prior means and variances, the smoothness hyperparameters can be chosen by trial-and-error simulations. Figure 4 also depicts four IRF draws from the multivariate Gaussian prior distribution with $\rho_{ij} = 0.9$ for all $(i, j)$, while Figure 5 shows four draws with $\rho_{ij} = 0.3$. The latter draws are more jagged and erratic than the former draws, and many economists would agree that the jaggedness of the $\rho_{ij} = 0.9$ draws are more in line with their prior information about the smoothness of the true IRFs in this application.

The flexible and graphical SVMA prior elicitation procedure contrasts with prior specification in standard SVARs. As discussed in Sections 2.1 and 6, common SVAR methods subtly constrain the implicit prior on IRFs. Bayesian analysis in the SVMA model is explicit about the prior on IRFs, and researchers can draw on standard Bayesian tools for conducting prior sensitivity analysis, cf. the Online Appendix.

Other priors. The multivariate Gaussian prior distribution is flexible and easy to visualize but other prior choices are feasible as well. My inference procedure does not rely on Gaussianity of the prior, as the simulation method in Section 3 only requires that the log prior density and its gradient are computable. Hence, it is straight-forward to impose a different prior correlation structure than (7), or to impose heavy-tailed or asymmetric prior distributions on certain impulse responses. Section 6 gives examples of priors that transparently impose well-known identifying restrictions from the SVAR literature.

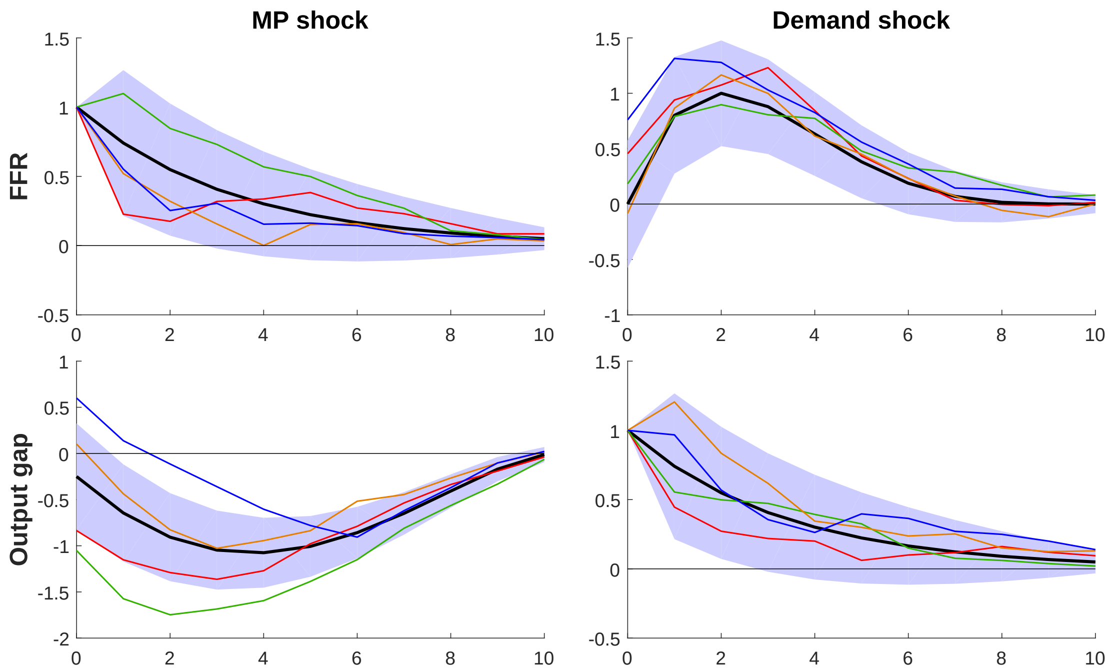


**Figure 4:** A choice of prior means (thick lines) and 90% prior confidence bands (shaded) for the four IRFs ($\Theta$) in the bivariate example in Figure 1. Brightly colored lines are four draws from the multivariate Gaussian prior distribution with these mean and variance parameters and a smoothness hyperparameter of $\rho_{ij} = 0.9$ for all $(i, j)$.

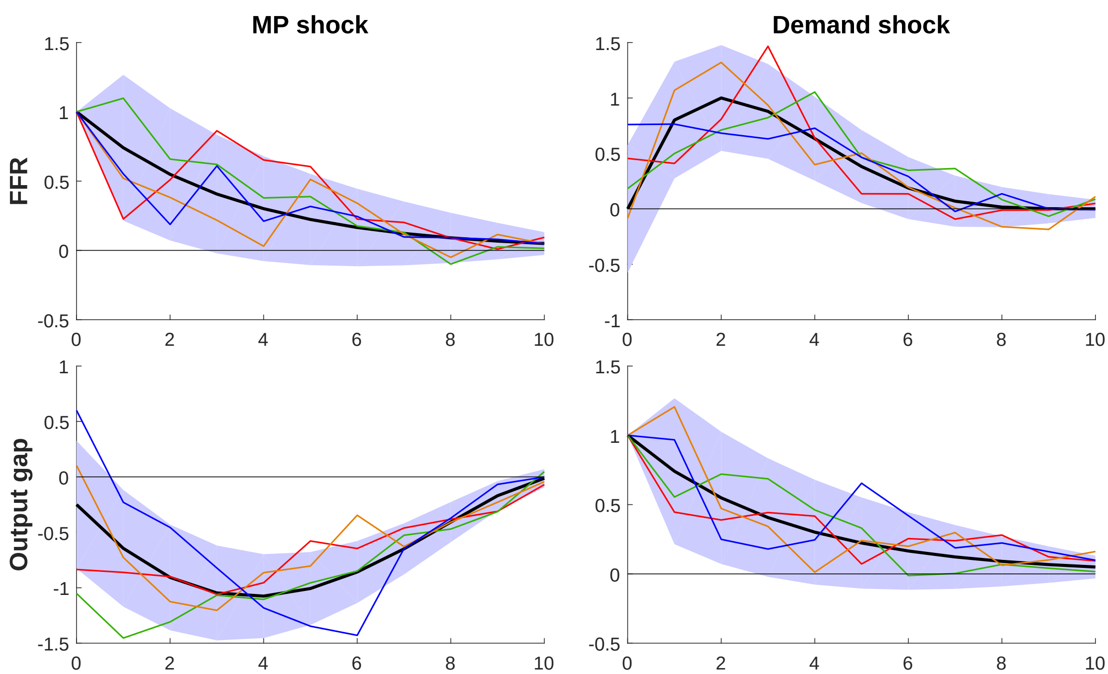


**Figure 5:** See caption for Figure 4. Here the smoothness parameter is $\rho_{ij} = 0.3$ for all $(i, j)$.

# 3 Bayesian computation

In this section I develop an algorithm to simulate from the posterior distribution of the IRFs. Because of the flexible and high-dimensional prior distribution placed on the IRFs, standard Markov Chain Monte Carlo (MCMC) methods are cumbersome.[21] I employ a Hamiltonian Monte Carlo algorithm that uses the Whittle (1953) likelihood approximation to speed up computations. The algorithm is fast, asymptotically efficient, and easy to apply, and it allows for both invertible and noninvertible IRFs.

I first define the posterior density of the structural parameters. Let $T$ be the sample size and $Y_T = (y_1', y_2', \dots, y_T')'$ the data vector. Denote the prior density for the SVMA parameters by $\pi_{\Theta,\sigma}(\Theta, \sigma)$. The likelihood function of the SVMA model (3) depends on the parameters $(\Theta, \sigma)$ only through the scaled impulse responses $\Psi = (\Psi_0, \Psi_1, \dots, \Psi_q)$, where $\Psi_\ell = \Theta_\ell \operatorname{diag}(\sigma)$ for $\ell = 0, 1, \dots, q$. Let $p_{Y|\Psi}(Y_T \mid \Psi(\Theta, \sigma))$ denote the likelihood function, where the notation indicates that $\Psi$ is a function of $(\Theta, \sigma)$. The posterior density is then

$$p_{\Theta,\sigma|Y}(\Theta, \sigma \mid Y_T) \propto p_{Y|\Psi}(Y_T \mid \Psi(\Theta, \sigma))\pi_{\Theta,\sigma}(\Theta, \sigma).$$

Hamiltonian Monte Carlo. To efficiently draw from the posterior distribution, I use a variant of MCMC known as Hamiltonian Monte Carlo (HMC). See Neal (2011) for an overview of HMC. By exploiting information contained in the gradient of the log posterior density to systematically explore the posterior distribution, HMC is known to outperform other generic MCMC methods in high-dimensional settings. In the SVMA model, the dimension of the full parameter vector is $n^2(q + 1)$, which can easily be well into the 100s in realistic applications. Nevertheless, the HMC algorithm has no trouble producing draws from the posterior of the SVMA parameters. I use the modified HMC algorithm by Hoffman & Gelman (2014), called the No-U-Turn Sampler (NUTS), which adaptively sets the HMC tuning parameters while still provably delivering draws from the posterior distribution.

As with other MCMC methods, the HMC algorithm delivers parameter draws from a Markov chain whose long-run distribution is the posterior distribution. After discarding a burn-in sample, the output of the HMC algorithm is a collection of parameter draws $(\Theta^{(1)}, \sigma^{(1)}), \dots, (\Theta^{(N)}, \sigma^{(N)})$, each of which is (very nearly) distributed according to the pos-

[21] Chib & Greenberg (1994) estimate univariate reduced-form Autoregressive Moving Average models by MCMC, but their algorithm is only effective in low-dimensional problems. Chan, Eisenstat & Koop (2015, see also references therein) perform Bayesian inference in possibly high-dimensional reduced-form VARMA models, but they impose statistical parameter normalizations that preclude structural estimation of IRFs.

terior distribution.[22] The number $N$ of draws is chosen by the user. The draws are not independent, and plots of the autocorrelation functions of the draws are useful for gauging the reduction in effective sample size relative to the ideal of i.i.d. sampling. In my experience, the proposed algorithm for the SVMA model yields autocorrelations that drop off to zero after only a few lags.

Likelihood, score and Whittle approximation. HMC requires that the log posterior density and its gradient can be computed quickly at any given parameter values. The gradient of the log posterior density equals the gradient of the log prior density plus the gradient of the log likelihood (the latter is henceforth referred to as the *score*). In most cases, such as with the Gaussian prior in Section 2.5, the log prior density and its gradient are easily computed. The log likelihood and the score are the bottlenecks. In the empirical study in the next section a full run of the HMC procedure requires 100,000s of evaluations of the likelihood and the score.

With Gaussian shocks (Assumption 2), the likelihood of the SVMA model (3) can be evaluated using the Kalman filter, but a faster alternative is to use the Whittle (1953) approximation to the likelihood of a stationary Gaussian process. See the Online Appendix for a description of the Kalman filter. Appendix A.3 shows that both the Whittle log likelihood and the Whittle score for the SVMA model can be calculated efficiently using the Fast Fourier Transform.[23] When the MA lag length $q$ is large, as in most applications, the Whittle likelihood is noticeably faster to compute than the exact likelihood, and massive computational savings arise from using the Whittle approximation to the score.

Numerical implementation. The HMC algorithm is easy to apply once the prior has been specified. I give further details on the Bayesian computations in the Online Appendix. As initial value for the HMC iterations I use a rough approximation to the posterior mode obtained using the characterization of the identified set in Appendix A.2. Matlab code for implementing the full inference procedure is available on my website, cf. Footnote 1. The Online Appendix illustrates the accuracy and rapid convergence of the Bayesian computations when applied to the bivariate model and prior in Figures 1 and 4, as well as to

[22] Gelman et al. (2013, Ch. 11) discuss methods for checking that the chain has converged.

[23] Hansen & Sargent (1981), Ito & Quah (1989), and Christiano & Vigfusson (2003) also employ the Whittle likelihood for SVMA models. Qu & Tkachenko (2012a,b) and Sala (2015) use the Whittle likelihood to perform approximate Bayesian inference on DSGE models, but their Random-Walk Metropolis-Hastings simulation algorithm is less efficient than HMC.

specifications in which the prior is centered far from the true parameter values.

Reweighting. The Online Appendix describes an optional reweighting step that translates the Whittle HMC draws into draws from the exact posterior $p_{\Theta,\sigma|Y}(\Theta, \sigma \mid Y_T)$. However, the asymptotic analysis in Section 5.3 shows that, at least for moderate lag lengths $q$, the reweighting step has negligible effect in large samples.

# 4 Application: News shocks and business cycles

I now use the SVMA method to infer the role of technological news shocks in the U.S. business cycle. Following the literature, I define a technological news shock to be a signal about future productivity increases. My prior on IRFs is informed by a conventional sticky-price DSGE model, without imposing the model restrictions dogmatically. The analysis finds overwhelming evidence of noninvertible IRFs in my specification. News shocks turn out to be relatively unimportant drivers of productivity and output growth, but more important for the real interest rate.

Technological news shocks have received great attention in the recent empirical and theoretical macro literature, but researchers have not yet reached a consensus on their importance, cf. the survey by Beaudry & Portier (2014). As explained in Section 2.3, theoretical macro models with news shocks often feature noninvertible IRFs, giving the SVMA method a distinct advantage over SVARs, as the latter assume away noninvertibility.

Specification and data. I employ a SVMA model with three observed variables and three unobserved shocks: Total factor productivity (TFP) growth, real gross domestic product (GDP) growth, and the real interest rate are assumed to be driven by a productivity shock, a technological news shock, and a monetary policy shock. I use quarterly data from 1954Q3–2007Q4, yielding sample size $T = 213$. I exclude data from 2008 to the present as my analysis ignores financial shocks.

The data set is detailed in the Online Appendix. TFP growth is obtained from Fernald (2014). The real interest rate equals the effective federal funds rate minus the contemporaneous GDP deflator inflation rate. The series are detrended using the kernel smoother in Stock & Watson (2012). I pick a MA lag length of $q = 16$ quarters based on two considerations. First, the Akaike Information Criterion (computed using the Whittle likelihood) selects $q = 13$. Second, the autocorrelation of the real interest rate equals 0.17 at lag 13 but

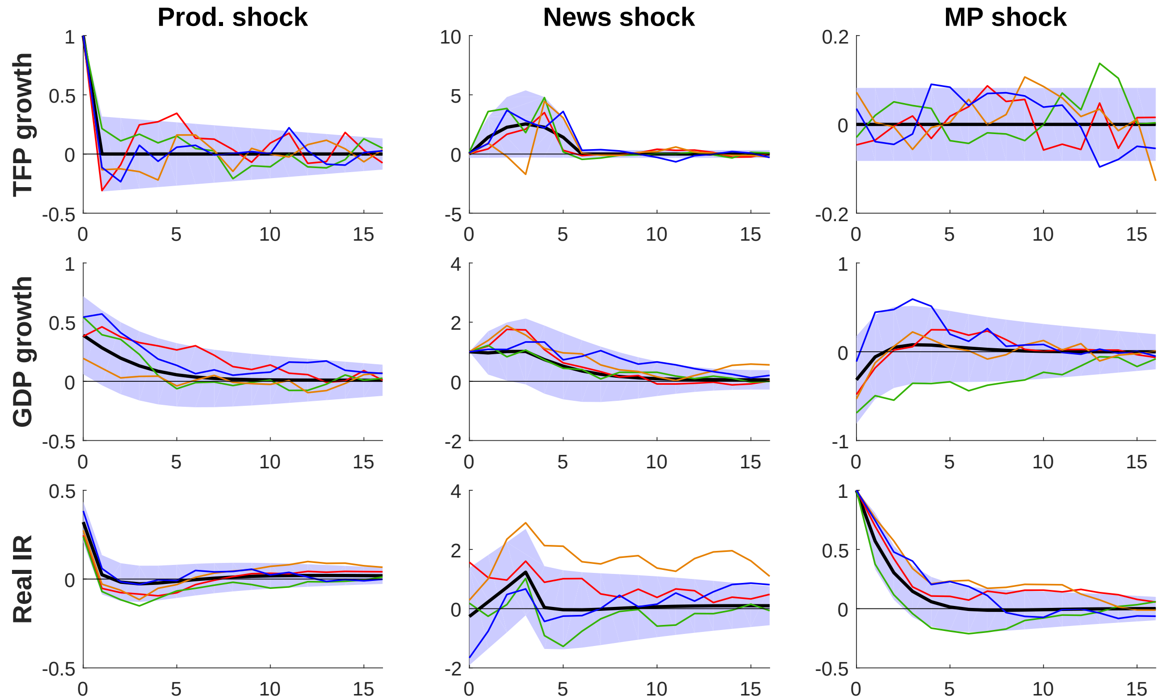


**Figure 6:** Prior means (thick lines), 90% prior confidence bands (shaded), and four random draws (brightly colored lines) from the prior for IRFs ($\Theta$), news shock application. The impact impulse response is normalized to 1 in each IRF along the diagonal of the figure.

is close to zero at lag 16.

Prior. The prior on the IRFs is of the multivariate Gaussian type introduced in Section 2.5, with hyperparameters informed by a conventional sticky-price DSGE model. The DSGE model is primarily used to guide the choice of prior means, and the model restrictions are not imposed dogmatically on the SVMA IRFs. Figure 6 plots the prior means and variances for the impulse responses, along with four draws from the joint prior distribution. The figure also shows the normalization that defines the scale of each shock.

The DSGE model used to inform the prior is the one developed by Sims (2012, Sec. 3). It is built around a standard New Keynesian structure with monopolistically competitive firms subject to a Calvo pricing friction, and the model adds capital accumulation, investment adjustment costs, habit formation, and interest rate smoothing. Within the DSGE model, the productivity and news shocks are, respectively, unanticipated and anticipated exogenous disturbances to the change in log TFP (cf. eq. 30–33 in Sims, 2012). The monetary policy shock is an unanticipated disturbance term in the Taylor rule (cf. eq. 35 in Sims, 2012). Detailed model assumptions and equilibrium conditions are described in Sims (2012, Sec.

3), but I repeat that I only use the DSGE model to guide the SVMA prior; the model restrictions are not imposed dogmatically.[24]

As prior means for the nine SVMA IRFs I use the corresponding IRFs implied by the log-linearized DSGE model, with one exception mentioned below.[25] I use the baseline calibration of Sims (2012, Table 1), which assumes that news shocks are correctly anticipated TFP increases taking effect three quarters into the future. Because I am particularly uncertain that an anticipation horizon of three quarters is correct, I modify the prior means for the impulse responses of TFP growth to the news shock: The prior means smoothly increase and then decrease over the interval $\ell \in [0, 6]$, with a maximum value at $\ell = 3$ equal to half the DSGE-implied impulse response.

The prior variances for the IRFs are chosen by combining information from economic intuition and DSGE calibration sensitivity experiments. For example, I adjust the prior variances for the IRFs so that the DSGE-implied IRFs mostly fall within the 90% prior bands when the anticipation horizon changes between nearby values. The 90% prior bands for the IRFs that correspond to the news shock are chosen quite large, and they mostly contain 0. In contrast, the prior bands corresponding to the monetary policy shock are narrower, expressing a strong belief that monetary policy shocks have a small effect on TFP growth but a persistent positive effect on the real interest rate. The prior bands for the effects of productivity shocks on GDP growth and on the real interest rate are fairly wide, since these IRFs should theoretically be sensitive to the degree of nominal rigidity as well as to the Federal Reserve's information set and policy rule.

The prior expresses a belief that the IRFs for GDP growth and the real interest rate are smooth, while those for TFP growth are less smooth. Specifically, I set $\rho_{1j} = 0.5$ and $\rho_{2j} = \rho_{3j} = 0.9$ for $j = 1, 2, 3$. These choices are consistent with standard calibrations of DSGE models. The ability to easily impose different degrees of prior smoothness across IRFs is unique to the SVMA approach; it would be much harder to achieve in a SVAR set-up.

The prior on the shock standard deviations is very diffuse. For each shock $j$, the prior mean $\mu_j^\sigma$ of $\log(\sigma_j)$ is set to $\log(0.5)$, while the prior standard deviation $\tau_j^\sigma$ is set to 2.[26]

[24] My approach differs from IRF matching (Rotemberg & Woodford, 1997). That method identifies a SVAR using exclusion restrictions, and then chooses the structural parameters of a DSGE model so that the DSGE-implied IRFs match the estimated SVAR IRFs. In my procedure, the DSGE model non-dogmatically informs the choice of prior on IRFs, but then the data is allowed to speak through a flexible SVMA model. Ingram & Whiteman (1994) and Del Negro & Schorfheide (2004) apply related ideas to VAR models.

[25] The DSGE-implied IRFs for the real interest rate use the same definition of this variable as in the construction of the data series. IRFs are computed using Dynare 4.4.3 (Adjemian et al., 2011).

[26] Unreported simulations show that the prior 5th and 95th percentiles of the FEVD (cf. (8)) are very

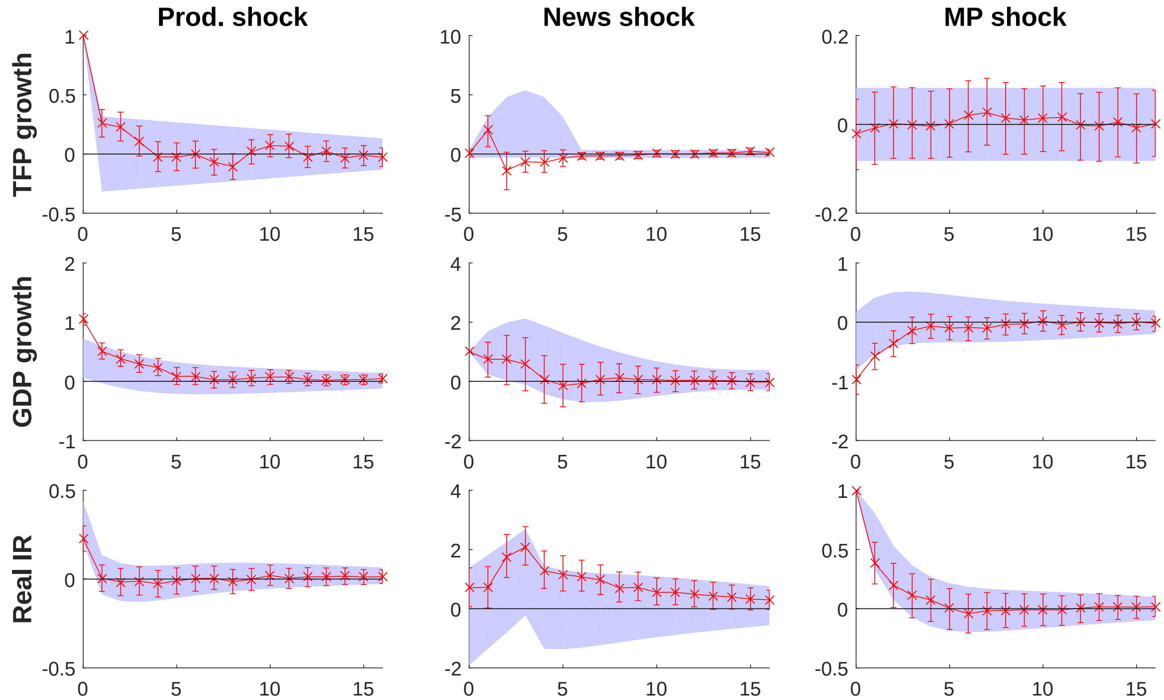


**Figure 7:** Summary of posterior IRF ($\Theta$) draws, news shock application. The plots show prior 90% confidence bands (shaded), posterior means (crosses), and posterior 5–95 percentile intervals (vertical bars).

These values should of course depend on the units of the observed series.

RESULTS. Given my prior, the data is informative about most of the IRFs. Figure 7 summarizes the posterior distribution of the IRFs. Figure 8 shows the posterior distribution of the forecast error variance decomposition (FEVD), defined as[27]

$$FEVD_{ij,\ell} = \frac{\text{Var}(\sum_{k=0}^{q} \Theta_{ij,k}\varepsilon_{j,t+\ell-k} \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)}{\text{Var}(y_{i,t+\ell} \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)} = \frac{\sum_{k=0}^{\ell} \Theta_{ij,k}^2 \sigma_j^2}{\sum_{b=1}^{n} \sum_{k=0}^{\ell} \Theta_{ib,k}^2 \sigma_b^2}. \tag{8}$$

$FEVD_{ij,\ell}$ is the fraction of the forecast error variance that would be eliminated if we knew all future realizations of shock $j$ when forming $\ell$-quarter-ahead forecasts of variable $i$ at time $t$ using knowledge of all shocks up to time $t-1$.

close to 0 and 1, respectively, for almost all $(i, j, \ell)$ combinations.

[27]The variances in the fraction are computed under the assumption that the shocks are serially and mutually independent. In the literature the FEVD is defined by conditioning on $(y_{t-1}, y_{t-2}, \dots)$ instead of $(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$. This distinction matters when the IRFs are noninvertible. Baumeister & Hamilton (2015a) conduct inference on the FEVD in a Bayesian SVAR, assuming invertibility.

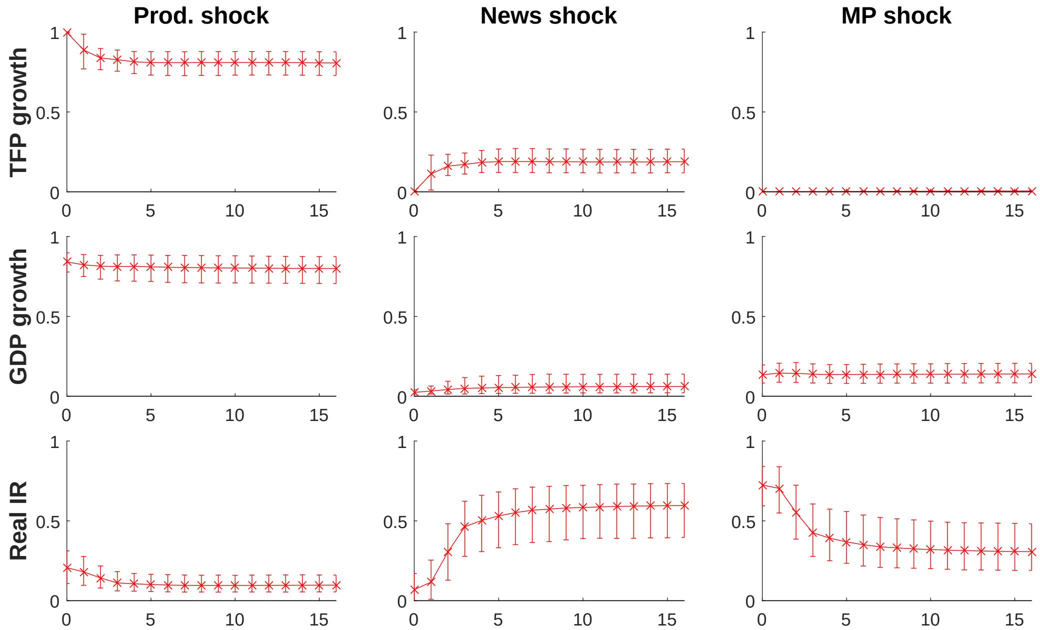


**Figure 8:** Summary of posterior draws of $FEVD_{ij,\ell}$ (8), news shock application. The figure shows posterior means (crosses) and posterior 5–95 percentile intervals (vertical bars). For each variable $i$ and each horizon $\ell$, the posterior means sum to 1 across the three shocks $j$.

The posterior means for several IRFs differ substantially from the prior means, and the posterior 90% intervals are narrower than the prior 90% bands. The effects of productivity and monetary policy shocks on TFP and GDP growth are especially precisely estimated. From the perspective of the prior beliefs, it is surprising to learn that the impact effect of productivity shocks on GDP growth is quite large, and the effect of monetary policy shocks on the real interest rate is not very persistent. The monetary policy shock has non-neutral (negative) effects on the level of GDP in the long run, even though the prior distribution for the cumulative response is centered around zero, cf. the Online Appendix.

The news shock is not an important driver of TFP and GDP growth but is important for explaining real interest rate movements. The IRF of TFP growth to the news shock indicates that future productivity increases are anticipated only one quarter ahead, and the increase is mostly reversed in the following quarters. According to the posterior, the long-run response of TFP to a news shock is unlikely to be substantially positive, implying that economic agents seldom correctly anticipate shifts in medium-run productivity levels. The news shock contributes little to the forecast error variance for TFP and GDP growth at all horizons. The monetary policy shock is only slightly more important for explaining GDP growth, while the productivity shock is much more important by these measures. However, the monetary

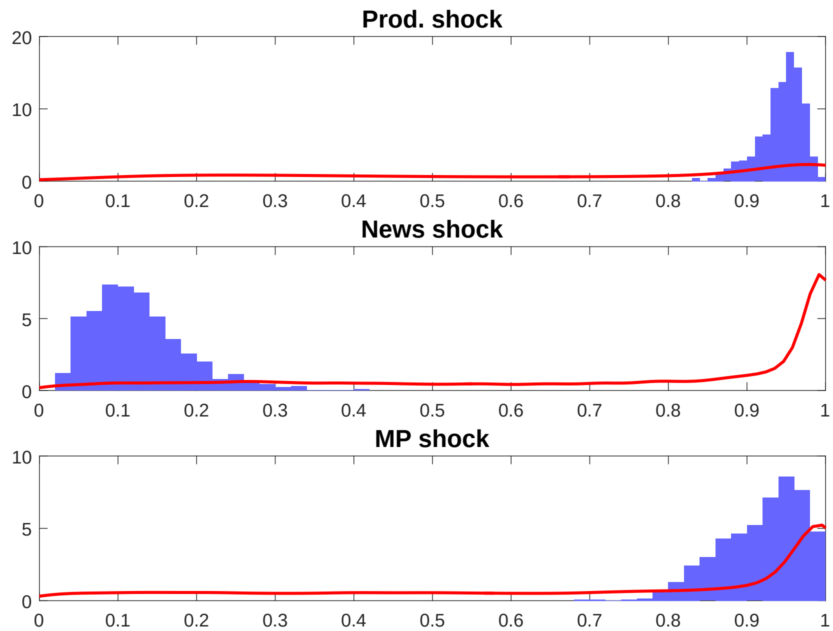


**Figure 9:** Histograms of posterior draws of the population $R^2$ values in regressions of each shock on current and 50 lagged values of the observed data, news shock application. Curves are kernel density estimates of the prior distribution of $R^2$s. Histograms and curves each integrate to 1.

policy shock is important for explaining short-run movements in the real interest rate, while the news shock dominates longer-run movements in this series.

The data and prior provide overwhelming evidence that the IRFs are noninvertible. In Figure 9 I report a continuous measure of invertibility suggested by Watson (1994, p. 2901) and Sims & Zha (2006, p. 243). For each posterior parameter draw I compute the $R^2$ in a population regression of each shock $\varepsilon_{j,t}$ on current and 50 lags of data $(y_t, y_{t-1}, \dots, y_{t-50})$, assuming i.i.d. Gaussian shocks.[28] This $R^2$ value should be essentially 1 for all shocks if the IRFs are invertible, by definition. Instead, Figure 9 shows a high posterior probability that the news shock $R^2$ is below 0.3, despite the prior putting most weight on values near 1.[29] The Online Appendix demonstrates that the noninvertibility is economically significant: The posterior distribution of the *invertible* IRFs that are closest (in a certain precise sense) to the actual IRFs is very different from the posterior distribution in Figure 7.

[28]Given the parameters, I run the Kalman filter in the Online Appendix forward for 51 periods on data that is identically zero (due to Gaussianity, conditional variances do not depend on realized data values). This yields a final updated state prediction variance matrix $\text{Var}(\text{diag}(\sigma)^{-1}\varepsilon_{51} \mid y_{51}, \dots, y_1)$ whose diagonal elements equal 1 minus the desired population $R^2$ values at the given parameters.

[29]Essentially no posterior IRF draws are exactly invertible; the prior probability is 0.06%.

Additional results and discussion. In the Online Appendix I plot the posterior distribution of the structural shocks across time, conduct prior sensitivity checks and posterior predictive checks of model validity, discuss how my empirical results relate to the literature, and verify that my method accurately recovers true IRFs on simulated data.

# 5 Asymptotic theory

To gain insight into how the data updates the prior information, I derive the asymptotic limit of the Bayesian posterior distribution from a frequentist point of view. I first derive general results on the frequentist asymptotics of Bayes procedures for a large class of partially identified models. Then I specialize to the SVMA model and show that, asymptotically, the role of the data is to pin down the true autocovariances, whereas all other information about IRFs comes from the prior. The asymptotics imply that the limiting form of the posterior is robust to violations of the assumption of Gaussian shocks and to the use of the Whittle likelihood in place of the exact likelihood.

## 5.1 General results for partially identified models

In this subsection I present a general result on the frequentist asymptotic limit of the Bayesian posterior distribution in partially identified models. Due to identification failure, the analysis is nonstandard, as the data does not dominate all aspects of the prior in large samples.

Consider a general model for which the data vector $Y_T$ is independent of the parameter of interest $\theta$, conditional on a second parameter $\Gamma$.[30] In other words, the likelihood function of the data $Y_T$ only depends on $\theta$ through $\Gamma$. This property holds for models with a partially identified parameter $\theta$, as explained in Poirier (1998). Because I will restrict attention to models in which the parameter $\Gamma$ is identified, I refer to $\Gamma$ as the reduced-form parameter, while $\theta$ is called the structural parameter. The parameter spaces for $\Gamma$ and $\theta$ are denoted $\Xi_\Gamma$ and $\Xi_\theta$, respectively, and these are assumed to be finite-dimensional Euclidean.

As an illustration, consider the SVMA model with data vector $Y_T = (y_1', \dots, y_T')'$. Let $\Gamma = (\Gamma(0), \dots, \Gamma(q))$ be the ACF of the observed time series, and let $\theta$ denote a single IRF, for example the IRF of the first variable to the first shock, i.e., $\theta = (\Theta_{11,0}, \dots, \Theta_{11,q})'$. I explain below why I focus on a single IRF. Since the distribution of the stationary Gaussian process $y_t$ only depends on $\theta$ through the ACF $\Gamma$, we have $Y_T \perp\!\!\!\perp \theta \mid \Gamma$.

[30] $T$ denotes the sample size, but the model does not have to be a time series model.

In any model satisfying $Y_T \perp\!\!\!\perp \theta \mid \Gamma$, the prior information about $\theta$ *conditional* on $\Gamma$ is not updated by the data $Y_T$, but the data is informative about $\Gamma$. Let $P_{\theta|Y}(\cdot \mid Y_T)$ denote the posterior probability measure for $\theta$ given data $Y_T$, and let $P_{\Gamma|Y}(\cdot \mid Y_T)$ denote the posterior measure for $\Gamma$. For any $\tilde{\Gamma} \in \Xi_\Gamma$, let $\Pi_{\theta|\Gamma}(\cdot \mid \tilde{\Gamma})$ denote the conditional prior measure for $\theta$ given $\Gamma$, evaluated at $\Gamma = \tilde{\Gamma}$. As in Moon & Schorfheide (2012, Sec. 3), decompose

$$P_{\theta|Y}(\mathcal{A} \mid Y_T) = \int_{\Xi_\Gamma} \Pi_{\theta|\Gamma}(\mathcal{A} \mid \Gamma) P_{\Gamma|Y}(d\Gamma \mid Y_T) \tag{9}$$

for any measurable set $\mathcal{A} \subset \Xi_\theta$. Let $\Gamma_0$ denote the true value of $\Gamma$. If the reduced-form parameter $\Gamma_0$ is identified, the posterior $P_{\Gamma|Y}(\cdot \mid Y_T)$ for $\Gamma$ will typically concentrate around $\Gamma_0$ in large samples, so that the posterior for $\theta$ is well approximated by $P_{\theta|Y}(\cdot \mid Y_T) \approx \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)$, the conditional prior for $\theta$ given $\Gamma$ at the true $\Gamma_0$.

The following lemma formalizes the intuition about the asymptotic limit of the posterior distribution for $\theta$. Define the $L_1$ norm $\|P\|_{L_1} = \sup_{|h|\leq 1} |\int h(x)P(dx)|$ on the space of signed measures, where $P$ is any signed measure and the supremum is over all scalar real-valued Borel measurable functions $h(\cdot)$ bounded in absolute value by 1.[31]

**Lemma 1.** *Let the posterior measure $P_{\theta|Y}(\cdot \mid Y_T)$ satisfy the decomposition (9). All stochastic limits below are taken under the true probability measure of the data. Assume:*

*(i) The map $\tilde{\Gamma} \mapsto \Pi_{\theta|\Gamma}(\theta \mid \tilde{\Gamma})$ is continuous at $\Gamma_0$ with respect to the $L_1$ norm $\|\cdot\|_{L_1}$.*[32]

*(ii) For any neighborhood $\mathcal{U}$ of $\Gamma_0$ in $\Xi_\Gamma$, $P_{\Gamma|Y}(\mathcal{U} \mid Y_T) \xrightarrow{p} 1$ as $T \to \infty$.*

*Then as $T \to \infty$,*

$$\|P_{\theta|Y}(\cdot \mid Y_T) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1} \xrightarrow{p} 0.$$

*If furthermore $\hat{\Gamma}$ is a consistent estimator of $\Gamma_0$, i.e., $\hat{\Gamma} \xrightarrow{p} \Gamma_0$, then*

$$\|P_{\theta|Y}(\cdot \mid Y_T) - \Pi_{\theta|\Gamma}(\cdot \mid \hat{\Gamma})\|_{L_1} \xrightarrow{p} 0.$$

[31]The $L_1$ distance $\|P_1 - P_2\|_{L_1}$ equals twice the total variation distance (TVD) between probability measures $P_1$ and $P_2$. Convergence in TVD implies convergence of Bayes point estimators under certain side conditions. In all results and proofs in this paper, the $L_1$ norm may be replaced by any (fixed) weaker norm for which the supremum is taken over a *subset* of measurable functions satisfying $|h(\cdot)| \leq 1$, e.g., the space of bounded Lipschitz functions.

[32]Denote the underlying probability sample space by $\Omega$, and let $\mathcal{B}_\theta$ be the Borel sigma-algebra on $\Xi_\theta$. Formally, assumption (i) requires the existence of a function $\varsigma \colon \mathcal{B}_\theta \times \Xi_\Gamma \to [0,1]$ such that $\{\varsigma(B, \Gamma(o))\}_{B\in\mathcal{B}_\theta,\, o\in\Omega}$ is a version of the regular conditional probability measure of $\theta$ given $\Gamma$, and such that $\|\varsigma(\cdot, \Gamma_k) - \varsigma(\cdot, \Gamma_0)\|_{L_1} \to 0$ as $k \to \infty$ for any sequence $\{\Gamma_k\}_{k\geq 1}$ satisfying $\Gamma_k \to \Gamma_0$ and $\Gamma_k \in \Xi_\Gamma$.

In addition to stating the explicit asymptotic form of the posterior distribution, Lemma 1 yields three main insights. First, the posterior for $\theta$ given the data does not collapse to a point asymptotically, a consequence of the lack of identification.[33] Second, the sampling uncertainty about the true reduced-form parameter $\Gamma_0$, which is identified in the sense of assumption (ii), is asymptotically negligible relative to the uncertainty about $\theta$ given knowledge of $\Gamma_0$. Third, in large samples, the way the data disciplines the prior information on $\theta$ is through the consistent estimator $\hat{\Gamma}$ of $\Gamma_0$.

Lemma 1 gives weaker and simpler conditions for result (ii) in Theorem 1 of Moon & Schorfheide (2012). Lipschitz continuity in $\Gamma$ of the conditional prior measure $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma)$ (their Assumption 2) is weakened to continuity, and the high-level assumption of asymptotic normality of the posterior for $\Gamma$ (their Assumption 1) is weakened to posterior consistency.

Assumption (i) invokes continuity with respect to $\Gamma$ of the conditional prior of $\theta$ given $\Gamma$. This assumption is satisfied in many models with partially identified parameters, if $\theta$ is chosen appropriately. The assumption is unlikely to be satisfied in other contexts. For example, if $\theta$ were identified because there existed a function mapping $\Gamma$ to $\theta$, and $\Gamma$ were identified, then assumption (i) could not be satisfied. More generally, assumption (i) will typically not be satisfied if the identified set for $\theta$ is a lower-dimensional subspace of $\Xi_\theta$.[34]

Assumption (ii) invokes posterior consistency for $\Gamma_0$, i.e., the posterior for the reduced-form parameter $\Gamma$ must concentrate on small neighborhoods of the true value $\Gamma_0$ in large samples. While assumption (i) is a condition on the prior, assumption (ii) may be viewed as a condition on the likelihood of the model, although assumption (ii) does require that the true reduced-form parameter $\Gamma_0$ is in the support of the marginal prior distribution for $\Gamma$. As long as the reduced-form parameter $\Gamma_0$ is identified, posterior consistency holds under weak regularity conditions, as discussed in the next subsection.

As the proof of Lemma 1 shows, the likelihood function used to calculate the posterior measure does not have to be correctly specified. That is, if $\tilde{\Gamma} \mapsto p_{Y|\Gamma}(Y_T \mid \tilde{\Gamma})$ denotes the likelihood function for $\Gamma$ used to compute the posterior $P_{\Gamma|Y}(\cdot \mid Y_T)$, then $p_{Y|\Gamma}(Y_T \mid \Gamma_0)$ need not be the true density of the data. As long as $P_{\Gamma|Y}(\cdot \mid Y_T)$ is a probability measure that satisfies the consistency assumption (ii), where the convergence in probability occurs under the true probability measure of the data, then the conclusion of the lemma follows.

[33] As emphasized by Gustafson (2015, pp. 35, 59–61), the Bayesian approach to partial identification acknowledges the role of prior information even in infinite samples, in contrast with "identified" models that ignore the bias arising from misspecification of the identifying restrictions.

[34] See Remarks 2 and 3, pp. 768–770, in Moon & Schorfheide (2012).

## 5.2 Posterior consistency for the autocovariance function

I now show that the posterior consistency assumption for the reduced-form parameter $\Gamma$ in Lemma 1 is satisfied in almost all stationary time series models for which $\Gamma$ can be chosen to be the ACF, such as the SVMA model. The result below supposes that the posterior measure for the ACF is computed using the Whittle likelihood, but posterior consistency obtains under general model misspecification.[35]

The only restrictions imposed on the underlying true data generating process are the following nonparametric stationarity and weak dependence assumptions.

**Assumption 3.** $\{y_t\}$ *is an* $n$*-dimensional time series satisfying the following assumptions. All limits and expectations below are taken under the true probability measure of the data.*

*(i)* $\{y_t\}$ *is a covariance stationary time series with mean zero.*

*(ii)* $\sum_{k=-\infty}^{\infty} \|\Gamma_0(k)\| < \infty$*, where the true ACF is defined by* $\Gamma_0(k) = E(y_{t+k}y_t')$*,* $k \in \mathbb{Z}$*.*

*(iii)* $\inf_{\omega\in[0,\pi)} \det\left(\sum_{k=-\infty}^{\infty} e^{-\imath k\omega}\Gamma_0(k)\right) > 0$*.*

*(iv) For any fixed integer* $k \geq 0$*,* $T^{-1}\sum_{t=k+1}^{T} y_t y_{t-k}' \xrightarrow{p} \Gamma_0(k)$ *as* $T \to \infty$*.*

The assumption imposes four weak conditions on $\{y_t\}$. First, the time series must be covariance stationary to ensure that the true ACF $\Gamma_0(\cdot)$ is well-defined (as usual, the mean-zero assumption can be easily relaxed). Second, the process is assumed to be weakly dependent, in the sense that the matrix ACF is summable, implying that the spectral density is well-defined. Third, the true spectral density must be uniformly non-singular, meaning that the process has full rank, is strictly nondeterministic, and has a positive definite ACF. Fourth, I assume the weak law of large numbers applies to the sample autocovariances.

To state the posterior consistency result, I first define the posterior measure. Let $\mathbb{T}_{n,q}$ be the space of ACFs for $n$-dimensional full-rank nondeterministic $q$-dependent processes, and let $p^W_{Y|\Gamma}(Y_T \mid \Gamma)$ denote the Whittle approximation to the likelihood of a stationary Gaussian process with ACF $\Gamma$ (these objects are defined in detail in Appendix A.4.1). Let $\Pi_\Gamma(\cdot)$ be a

[35] In the case of i.i.d. data, posterior consistency in misspecified models has been investigated in detail, see references in Ramamoorthi, Sriram & Martin (2015). Shalizi (2009) proves posterior consistency under misspecification with dependent data, placing high-level assumptions on the prior and likelihood. Müller (2013) discusses decision theoretic properties of Bayes estimators when the model is mispecified. Tamaki (2008) derives an asymptotic approximation to the Whittle posterior under correct specification.

prior measure on $\mathbb{T}_{n,q}$. The associated Whittle posterior measure for $\{\Gamma_0(k)\}_{0\leq k\leq q}$ is

$$P^W_{\Gamma|Y}(\mathcal{A} \mid Y_T) = \frac{\int_{\mathcal{A}} p^W_{Y|\Gamma}(Y_T \mid \Gamma)\Pi_\Gamma(d\Gamma)}{\int_{\mathbb{T}_{n,q}} p^W_{Y|\Gamma}(Y_T \mid \Gamma)\Pi_\Gamma(d\Gamma)}, \tag{10}$$

for any measurable subset $\mathcal{A}$ of $\mathbb{T}_{n,q}$.

**Proposition 1.** *Let Assumption 3 hold. Assume that $\{\Gamma_0(k)\}_{0\leq k\leq q}$ is in the support of $\Pi_\Gamma(\cdot)$. Then the Whittle posterior for $\{\Gamma_0(k)\}_{0\leq k\leq q}$ is consistent, i.e., for any neighborhood $\mathcal{U}$ of $\{\Gamma_0(k)\}_{0\leq k\leq q}$ in $\mathbb{T}_{n,q}$, we have $P^W_{\Gamma|Y}(\mathcal{U} \mid Y_T) \xrightarrow{p} 1$ as $T \to \infty$, under the true probability measure of the data.*

The SVMA model with Gaussian shocks is an example of a model with a stationary Gaussian and $q$-dependent likelihood. Hence, when applied to the SVMA model, Proposition 1 states that if the prior measure on the SVMA parameters induces a prior measure on $\Gamma$ that has the true ACF $\{\Gamma_0(k)\}_{0\leq k\leq q}$ in its support, then the model-implied Whittle posterior for $\Gamma$ pins down the true ACF in large samples. Proposition 1 places no restrictions on the prior $\Pi_\Gamma(\cdot)$ on the ACF, except that the true ACF $\Gamma_0$ lies in its support.

Posterior consistency for the true ACF (up to lag $q$) holds even for time series that are not Gaussian or $q$-dependent. This is true even though the measure $P^W_{\Gamma|Y}(\mathcal{A} \mid Y_T)$ is computed using the Whittle likelihood and therefore exploits the working assumption that the data is Gaussian and $q$-dependent. The only restrictions placed on the true distribution of the data are the stationarity and weak dependence conditions in Assumption 3. Proposition 1 is silent about posterior inference on autocovariances at lags higher than $q$.

## 5.3 Limiting posterior distribution in the SVMA model

I finally specialize the general asymptotic results from the previous subsections to the SVMA model with a non-dogmatic prior on IRFs. The asymptotics allow for noninvertibility and non-Gaussian structural shocks. The frequentist large-sample approximation to the Bayesian posterior shows that the role of the data is to pin down the true autocovariances of the data, which in turn pins down the reduced-form (Wold) IRFs, while all other information about the structural IRFs comes from the prior. I also argue that the limiting form of the posterior is the same whether the Whittle likelihood or the exact likelihood is used.

Set-up and main result. To map the SVMA model into the general framework, let $\theta$ denote the IRFs and shock standard deviation corresponding to the first shock, and let $\Gamma$

denote the ACF of the data: $\theta = (\{\Theta_{i1,\ell}\}_{1\leq i\leq n,\, 0\leq \ell\leq q}, \sigma_1)$ and $\Gamma = (\Gamma(0), \dots, \Gamma(k))$. I now apply Lemma 1 and Proposition 1 to the SVMA model, which gives a simple description of the limiting form of the Whittle posterior $P^W_{\theta|Y}(\cdot \mid Y_T)$ for all the structural parameters pertaining to the first shock. This analysis of course applies to each of the other shocks.

I choose $\theta$ to be the IRFs and shock standard deviation corresponding to a single shock in order to satisfy the prior continuity assumption in Lemma 1. In the SVMA model,

$$\Gamma(k) = \sigma_1^2 \sum_{\ell=0}^{q-k} \Theta_{:1,\ell+k}\Theta'_{:1,\ell} + \sum_{j=2}^{n} \sigma_j^2 \sum_{\ell=0}^{q-k} \Theta_{:j,\ell+k}\Theta'_{:j,\ell}, \quad k = 0, 1, \dots, q, \tag{11}$$

where $\Theta_{:j,\ell} = (\Theta_{1j,\ell}, \dots, \Theta_{nj,\ell})'$. If $\theta = (\{\Theta_{i1,\ell}\}_{1\leq i\leq n,\, 0\leq \ell\leq q}, \sigma_1)$ and there are two or more shocks ($n \geq 2$), then the above equations for $k = 0, 1, \dots, q$ are of the form $\Gamma = G(\theta) + U$, where $G(\cdot)$ is a matrix-valued function and $U$ is a function only of structural parameters pertaining to shocks $j \geq 2$. $\theta$ and $U$ are *a priori* independent provided that the $n^2$ IRFs and $n$ shock standard deviations are *a priori* mutually independent (for example, the multivariate Gaussian prior in Section 2.5 imposes such independence). In this case, the reduced-form parameter $\Gamma$ equals a function of the structural parameter $\theta$ plus *a priori* independent "noise" $U$. If the prior on the IRFs is non-dogmatic so that $U$ has full support, we can expect the conditional prior distribution of $\theta$ given $\Gamma$ to be continuous in $\Gamma$.[36]

On the other hand, the conditional prior distribution for $\theta$ given $\Gamma$ would not be continuous in $\Gamma$ if I had picked $\theta$ to be *all* IRFs and shock standard deviations. If $\theta = (\Theta, \sigma)$, then $\Gamma$ would equal a deterministic function of $\theta$, cf. (11), and so continuity of the conditional prior $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma)$ would not obtain. Hence, Lemma 1 is not useful for deriving the limit of the *joint* posterior of all structural parameters of the SVMA model.

The main proposition below states the limiting form of the Whittle posterior under general choices for the prior on IRFs and shock standard deviations. That is, I do not assume the multivariate Gaussian prior from Section 2.5. I also do not restrict the prior to the region of invertible IRFs, unlike the implicit priors used in SVAR analysis. Let $\Pi_{\Theta,\sigma}(\cdot)$ denote any prior measure for $(\Theta, \sigma)$ on the space $\Xi_\Theta \times \Xi_\sigma$. Through equation (6), this prior induces a joint prior measure $\Pi_{\Theta,\sigma,\Gamma}(\cdot)$ on $(\Theta, \sigma, \Gamma)$, which in turn implies marginal prior measures $\Pi_\theta(\cdot)$ and $\Pi_\Gamma(\cdot)$ for $\theta$ and $\Gamma$ as well as the conditional prior measure $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma)$ for $\theta$ given $\Gamma$. Let $P^W_{\theta|Y}(\cdot \mid Y_T)$ denote the Whittle posterior measure for $\theta$ computed using the Whittle SVMA likelihood, cf. Section 3, and the prior $\Pi_{\Theta,\sigma}(\cdot)$.

[36] This paragraph is inspired by Remark 3, pp. 769–770, in Moon & Schorfheide (2012).

**Proposition 2.** *Let the data $Y_T = (y_1', \dots, y_T')'$ be generated from a time series $\{y_t\}$ satisfying Assumption 3 (but not necessarily Assumptions 1 and 2). Assume that the prior $\Pi_{\Theta,\sigma}(\cdot)$ for $(\Theta, \sigma)$ has full support on $\Xi_\Theta \times \Xi_\sigma$. If the induced conditional prior $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma)$ satisfies the continuity assumption (i) of Lemma 1, then the Whittle posterior satisfies*

$$\|P^W_{\theta|Y}(\cdot \mid Y_T) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1} \xrightarrow{p} 0,$$

*as $T \to \infty$ under the true probability measure of the data. The above convergence also holds with $\Gamma_0$ replaced by $\hat{\Gamma} = \{\hat{\Gamma}(k)\}_{0 \le k \le q}$, the sample ACF.*

Continuity of the conditional prior $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma)$ is stated as a high-level assumption in Proposition 2. I conjecture that prior continuity holds for the multivariate Gaussian prior introduced in Section 2.5, for the reasons discussed below equation (11).

An important caveat on the results in this subsection is that the MA lag length $q$ is considered fixed as the sample size $T$ tends to infinity. In applications where $q$ is large relative to $T$, i.e., when the data is very persistent, these asymptotics may not be a good guide to the finite-sample behavior of the posterior. Nevertheless, the fixed-$q$ asymptotics do shed light on the interplay between the SVMA model, the prior, and the data.[37]

How the data updates the prior. According to Proposition 2, the posterior for the structural parameters $\theta$ does not collapse to a point asymptotically, but the data does pin down the true ACF $\Gamma_0$. Equivalently, the data reveals the true *reduced-form* IRFs and innovation variance matrix, or more precisely, reveals the Wold representation of the observed time series $y_t$ (Hannan, 1970, Thm. 2″, p. 158). Due to the under-identification of the SVMA model, many different *structural* IRFs are observationally equivalent with the Wold IRFs, cf. Appendix A.2. In large samples, the prior is the only source of information able to discriminate between different structural IRFs that are consistent with the true ACF.

Unlike SVARs, the SVMA approach does not infer long-horizon IRFs only from short-run dynamics of the data. In large samples the SVMA posterior depends on the data through the empirical autocovariances $\hat{\Gamma}$ out to lag $q$. Inference about long-horizon impulse responses is informed by the empirical autocovariances at the same long horizons (as well as other horizons). In contrast, most SVAR estimation procedures extrapolate long-horizon IRFs from the first few empirical autocorrelations of the data (Jordà, 2005).

[37] I conjecture that my results can be extended to the asymptotic embedding $q = q(T) = O(T^\nu)$, for appropriate $\nu > 0$ and under additional nonparametric conditions.

Proposition 2 shows to what extent the data can falsify the prior. The data indicates whether the induced prior $\Pi_\Gamma(\cdot)$ on the ACF is at odds with the true ACF $\Gamma_0$. For example, if the prior distribution on IRFs imposes a strong (but non-dogmatic) belief that $\{y_t\}$ is very persistent, but the actual data generating process is not persistent, the posterior will in large samples put most mass on IRFs that imply low persistence, as illustrated in the Online Appendix. On the other hand, if the prior on IRFs is tightly concentrated around parameters $(\Theta, \sigma)$ that lie in the identified set $\mathcal{S}(\Gamma_0)$, cf. Section 2.4, then the posterior also concentrates around $(\Theta, \sigma)$, regardless of how close $(\Theta, \sigma)$ are to the true structural parameters.

Robustness to misspecified likelihood. Proposition 2 states that the posterior measure, computed using the Whittle likelihood and thus under the working assumption of a Gaussian SVMA model, converges to $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)$ regardless of whether the Gaussian SVMA model is correctly specified.[38] The only restrictions on the true data generating process are the stationarity and weak dependence conditions in Assumption 3. Of course, the IRF parameters only have a structural economic interpretation if the basic SVMA model holds. In this case, the ACF has the form (6), so the conditional prior $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)$ imposes valid restrictions on the structural parameters. Thus, under Assumptions 1 and 3, the large-sample shape of the Whittle SVMA posterior provides valid information about $\theta$ even when the shocks are non-Gaussian or heteroskedastic (i.e., $E(\varepsilon_{j,t}^2 \mid \{\varepsilon_s\}_{s<t})$ is non-constant).[39]

The asymptotic robustness to non-Gaussianity is a consequence of the negligible importance of the uncertainty surrounding estimation of the true ACF $\Gamma_0$. As in the general Lemma 1, the latter uncertainty gets dominated in large samples by the conditional prior uncertainty about the structural parameters $\theta$ given knowledge of $\Gamma_0$. Because the sampling distribution of any efficient estimator of $\Gamma_0$ in general depends on fourth moments of the data, it is sensitive to departures from Gaussianity, but this sensitivity does not matter for the first-order asymptotic limit of the posterior for the partially identified parameter $\theta$.

My results do not and cannot imply that Bayesian inference based on the Gaussian SVMA model is asymptotically equivalent to *optimal* Bayesian inference under non-Gaussian shocks. If the SVMA likelihood were computed under the assumption that the structural shocks $\varepsilon_t$ are i.i.d. Student-t distributed, say, then the asymptotic limit of the posterior would differ

[38] Baumeister & Hamilton (2015b) derive an analogous result for a Bayesian SVAR model with a particular family of prior distributions and assuming invertibility.

[39] Standard arguments show that Assumption 1 implies Assumption 3 under two additional conditions: The true polynomial $\Theta(z)$ cannot have any roots exactly on the unit circle (but the true IRFs may be invertible or noninvertible), and the shocks $\varepsilon_t$ must have enough moments to ensure consistency of $\hat{\Gamma}$.

from $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)$. Indeed, if the shocks are known to be non-Gaussian, then higher-order cumulants of the data have identifying power, the empirical ACF does not constitute an asymptotically sufficient statistic for the IRFs, and it may no longer be the case that every invertible set of IRFs can be matched with an observationally equivalent set of noninvertible IRFs (Lanne & Saikkonen, 2013; Gospodinov & Ng, 2015).

However, Bayesian inference based on non-Gaussian shocks is less robust than Gaussian inference. Intuitively, while the expectation of the Gaussian or Whittle (quasi) log likelihood function depends only on second moments of the data, the expectation of a non-Gaussian log likelihood function generally depends also on higher moments. Hence, Bayesian inference computed under non-Gaussian shocks is misleading asymptotically if a failure of the distributional assumptions causes misspecification of higher-order moments, even if second moments are correctly specified.

Proposition 2 also implies that the error incurred in using the Whittle approximation to the SVMA likelihood is negligible in large samples: The data pins down the true ACF asymptotically even when the Whittle approximation is used. This is true whether or not the data is generated by a Gaussian SVMA model, as long as Assumption 3 holds.

# 6 Comparison with SVAR methods

This section shows that standard SVAR identifying restrictions can be transparently imposed through specific prior choices in the SVMA model, if desired.[40]

The most popular identifying restrictions in the literature are exclusion (i.e., zero) restrictions on short-run (i.e., impact) impulse responses: $\Theta_{ij,0} = 0$ for certain pairs $(i, j)$. These short-run exclusion restrictions include so-called "recursive" or "Cholesky" orderings, in which the $\Theta_0$ matrix is assumed triangular. Exclusion restrictions on impulse responses (at horizon 0 or higher) can be incorporated in the SVMA framework by simply setting the corresponding $\Theta_{ij,\ell}$ parameters equal to zero and dropping them from the parameter vector. Prior elicitation and posterior computation for the remaining parameters are unaffected.

Another popular type of identifying restrictions are exclusion restrictions on long-run (i.e., cumulative) impulse responses: $\sum_{\ell=0}^{q} \Theta_{ij,\ell} = 0$ for certain pairs $(i, j)$. Long-run exclusion restrictions can be accommodated in the SVMA model by restricting $\Theta_{ij,q} = -\sum_{\ell=0}^{q-1} \Theta_{ij,\ell}$ when evaluating the likelihood. The first $q$ impulse responses $(\Theta_{ij,0}, \dots, \Theta_{ij,q-1})$ are treated as free parameters whose prior must be specified by the researcher. When evaluating the

[40]The online appendix to Barnichon & Matthes (2016) discusses dogmatic SVMA identification restrictions.

score in the HMC procedure, cf. Section 3, the chain rule must be used to incorporate the effect that a change in $\Theta_{ij,\ell}$ ($\ell < q$) has on the implied value for $\Theta_{ij,q}$.

Short- or long-run exclusion restrictions are special cases of linear restrictions on the IRF parameters. Suppose we have prior information that $C\,\text{vec}(\Theta) = d$, where $C$ is a known full-rank matrix and $d$ is a known vector.[41] Let $C^{\perp}$ be a matrix such that $(C', C^{\perp})$ is a square invertible matrix and $CC^{\perp} = 0$. We can then reparametrize $\text{vec}(\Theta) = C^{\perp}\psi + C'(CC')^{-1}d$, where $\psi$ is an unrestricted vector. Given a prior for $\psi$,[42] posterior inference in the SVMA model can be carried out as in Section 3, except that $\Theta$ is treated as a known linear function of the free parameters $\psi$. Again, the chain rule provides the score with respect to $\psi$.

The preceding discussion dealt with *dogmatic* prior restrictions that impose exclusion restrictions with 100% prior certainty, but in many cases *non-dogmatic* restrictions are more credible.[43] Multivariate Gaussian priors can easily handle non-dogmatic prior restrictions. A prior belief that the impulse response $\Theta_{ij,\ell}$ is close to zero with high probability is imposed by choosing prior mean $\mu_{ij,\ell} = 0$ along with a small value for the prior variance $\tau_{ij,\ell}^2$ (see the notation in Section 2.5). To impose a prior belief that the long-run impulse response $\sum_{\ell=0}^{q} \Theta_{ij,\ell}$ is close to zero with high probability, imagine that $\Theta_{ij,q} = -\sum_{\ell=0}^{q-1} \Theta_{ij,\ell} + \nu_{ij}$, where $\nu_{ij}$ is mean-zero independent Gaussian noise with a small variance. Given a choice of Gaussian prior for the first $q$ impulse responses $(\Theta_{ij,0}, \dots, \Theta_{ij,q-1})$, this relationship implies a prior on the entire IRF $(\Theta_{ij,0}, \dots, \Theta_{ij,q})$. These considerations only concern the functional form of the prior density; evaluation of the likelihood and score proceeds as in Section 3.

Many SVAR papers exploit sign restrictions on impulse responses: $\Theta_{ij,\ell} \geq 0$ or $\Theta_{ij,\ell} \leq 0$ for certain triplets $(i, j, \ell)$. Dogmatic sign restrictions can be imposed in the SVMA framework by restricting the IRF parameter space $\Xi_{\Theta}$ to the subspace where the inequality constraints hold (e.g., using reparametrization; see also Neal, 2011, Sec. 5.1). The prior distribution for the impulse responses in question can be chosen to be diffuse on the relevant subspace, if desired (e.g., truncated normal with large variance).[44]

However, researchers often have more prior information about impulse responses than just their signs, and this can be exploited in the SVMA approach. For example, extremely

[41] These restrictions should include the normalizations $\Theta_{i_j j,0} = 1$ for $j = 1, \dots, n$.

[42] Given a preliminary prior for $\Theta$, one could obtain a prior for $\psi$ from the relationship $\psi = (C^{\perp\prime}C^{\perp})^{-1}C^{\perp\prime}\{\text{vec}(\Theta) - C'(CC')^{-1}d\}$. In general, subtle issues (the Borel paradox) arise when a restricted prior is obtained from an unrestricted one (Drèze & Richard, 1983, Sec. 1.3).

[43] The distinction between dogmatic (exact) and non-dogmatic ("stochastic") identifying restrictions is familiar from the Bayesian literature on simultaneous equation models (Drèze & Richard, 1983).

[44] Poirier (1998, Sec. 4) warns against entirely flat priors in partially identified models.

large values for some impulse responses can often be ruled out *a priori*.[45] The Gaussian prior in Section 2.5 is capable of expressing a strong but non-dogmatic prior belief that certain impulse responses have certain signs, while expressing disbelief in extreme values. In some applications, a heavy-tailed or skewed prior distribution may be more appropriate.

Although computationally attractive, the Uhlig (2005) inference procedure for sign-identified SVARs is less transparent than the SVMA approach. Because Uhlig's prior is mainly chosen for computational convenience, it is not very flexible. Furthermore, Baumeister & Hamilton (2015b) show that the Uhlig (2005) procedure imposes unintended and unacknowledged prior information on IRFs in addition to the acknowledged sign restrictions.[46] In contrast, the SVMA prior is flexible and easy to visualize transparently.

The SVMA approach can exploit the identifying power of external instruments. An external instrument is an observed variable $z_t$ that is correlated with one of the structural shocks but uncorrelated with the other shocks (Stock & Watson, 2008, 2012; Mertens & Ravn, 2013). Such an instrument can be incorporated in the analysis by adding $z_t$ to the vector $y_t$ of observed variables. Suppose we add it as the first element ($i = 1$), and that $z_t$ is an instrument for the first structural shock ($j = 1$). The properties of the external instrument then imply that we have a strong prior belief that $\Theta_{1j,0}$ is (close to) zero for $j = 2, 3, \dots, n$. We may also have reason to believe that $\Theta_{1j,\ell} \approx 0$ for $\ell \geq 1$.

Finally, the SVMA IRFs can be restricted to be invertible, if desired, by rejecting posterior draws outside the invertible region $\{\Theta\colon \det(\sum_{\ell=0}^{q} \Theta_\ell z^\ell) \neq 0 \; \forall\, z \in \mathbb{C} \text{ s.t. } |z| < 1\}$.[47]

# 7 Topics for future research

I conclude by listing some avenues for future research. First, it would be interesting to investigate whether prior elicitation and posterior computation can be simplified if the researcher only cares about the effects of a single shock (while allowing for other shocks). Second, the SVMA model can allow for an infinite lag length $q$ if one specifies a parametric structure for the tail of the IRFs. Third, most procedures and formulas in this paper extend

[45]See the SVAR analyses by Kilian & Murphy (2012) and Baumeister & Hamilton (2015c). While these papers focus on prior information about impact impulse responses, the SVMA approach facilitates imposing information about longer-horizon responses.

[46]Giacomini & Kitagawa (2015) develop a robust Bayes SVAR approach that imposes dogmatic exclusion and sign restrictions without imposing any other identifying restrictions. My SVMA approach instead seeks to allow for as many types of prior information as possible.

[47]$\det(\Theta_0) = 0$ implies noninvertibility. Otherwise, the roots of $\det(\sum_{\ell=0}^{q} \Theta_\ell z^\ell)$ equal the roots of $\det(I_n + \sum_{\ell=1}^{q} \Theta_0^{-1} \Theta_\ell z^\ell)$, which equal the reciprocals of the eigenvalues of the polynomial's companion matrix.

directly to models with more shocks than observed variables, although the identification analysis would differ. Fourth, while I have shown Gaussianity-based posterior inference to be asymptotically robust, it would be interesting to explore the computational feasibility and robustness properties of explicitly incorporating nonstationarity, nonlinearities, time-varying parameters, non-Gaussian shocks, or stochastic volatility.[48] Finally, the posterior simulation algorithm can likely be improved in cases where an extremely diffuse prior causes the posterior distribution to be very multimodal.

# A Appendix

## A.1 Notation

$I_n$ is the $n \times n$ identity matrix. $\imath$ is the imaginary unit. For a vector $a$, $\text{diag}(a)$ denotes the diagonal matrix with the elements of $a$ along the diagonal in order. For a square matrix $A$, $\text{tr}(A)$ and $\det(A)$ are the trace and determinant, and $\text{diag}(A)$ is the vector consisting of the diagonal elements in order. For any matrix $B$, $B'$ denotes the matrix transpose, $\bar{B}$ denotes the elementwise complex conjugate, $B^* = \bar{B}'$ is the complex conjugate transpose, $\text{Re}(B)$ is the real part of $B$, $\|B\| = \sqrt{\text{tr}(B^* B)}$ is the Frobenius norm, and $\text{vec}(B)$ is the columnwise vectorization. An $n \times n$ matrix $Q$ is unitary if $QQ^* = I_n$, and a real unitary matrix is orthogonal. Independence of random variables $X$ and $Y$ conditional on $Z$ is denoted by $X \perp\!\!\!\perp Y \mid Z$. $\mathcal{K}^c$ denotes the complement of a set $\mathcal{K}$.

## A.2 Constructive characterization of the identified set

The result below applies the analysis of Lippi & Reichlin (1994) to the SVMA model; see also Hansen & Sargent (1981) and Komunjer & Ng (2011). I identify a set of IRFs $\Theta = (\Theta_0, \dots, \Theta_q)$ with the matrix polynomial $\Theta(z) = \sum_{\ell=0}^{q} \Theta_\ell z^\ell$, and I use the notation $\Theta$ and $\Theta(z)$ interchangeably where appropriate. The proposition states that if we start with a set of IRFs $\Theta(z)$ in the identified set, then we can obtain all other sets of IRFs in the identified set by orthogonally rotating $\Theta(z)$ and/or by "flipping the roots" of $\Theta(z)$. Only a finite sequence of such operations is necessary to jump between any two elements of the identified set.

[48] For cointegrated time series, an analog of the SVMA approach is to do Bayesian inference on the parameters of the Granger representation.

**Proposition 3.** *Let* $\{\Gamma(k)\}_{0\leq k\leq q}$ *be an arbitrary ACF. Pick an arbitrary* $(\Theta,\sigma)\in\mathcal{S}(\Gamma)$ *satisfying* $\det(\Theta(0))\neq 0$.[49] *Define* $\Psi(z)=\Theta(z)\,\mathrm{diag}(\sigma)$.

*Construct a matrix polynomial* $\check{\Psi}(z)$ *in either of the following two ways:*

*(i) Set* $\check{\Psi}(z)=\Psi(z)Q$*, where* $Q$ *is an arbitrary orthogonal* $n\times n$ *matrix.*

*(ii) Let* $\gamma_1,\dots,\gamma_r$ *(*$r\leq nq$*) denote the roots of the polynomial* $\det(\Psi(z))$*. Pick an arbitrary positive integer* $k\leq r$*. Let* $\eta\in\mathbb{C}^n$ *be a vector such that* $\Psi(\gamma_k)\eta=0$ *(such a vector exists because* $\det(\Psi(\gamma_k))=0$*). Let* $Q$ *be a unitary matrix whose first column is proportional to* $\eta$ *(if* $\gamma_k$ *is real, choose* $Q$ *to be a real orthogonal matrix). All elements of the first column of the matrix polynomial* $\Psi(z)Q$ *then contain the factor* $(z-\gamma_k)$*. In each element of the first column, replace the factor* $(z-\gamma_k)$ *with* $(1-\overline{\gamma_k}z)$*. Call the resulting matrix polynomial* $\check{\Psi}(z)$*. If* $\gamma_k$ *is real, skip the next paragraph.*

*If* $\gamma_k$ *is not real, let* $\tilde{\eta}\in\mathbb{C}^n$ *be a vector such that* $\check{\Psi}(\overline{\gamma_k})\tilde{\eta}=0$*, and let* $\tilde{Q}$ *be a unitary matrix whose first column is proportional to* $\tilde{\eta}$*. All elements of the first column of* $\check{\Psi}(z)\tilde{Q}$ *then contain the factor* $(z-\overline{\gamma_k})$*. In each element of the first column, replace the factor* $(z-\overline{\gamma_k})$ *with* $(1-\gamma_k z)$*. Call the resulting matrix polynomial* $\tilde{\Psi}(z)$*. The matrix* $\tilde{\Psi}(0)\overline{\tilde{\Psi}(0)}^{*}$ *is real, symmetric, and positive definite, so let* $J$ *be its* $n\times n$ *Cholesky factor:* $JJ'=\tilde{\Psi}(0)\overline{\tilde{\Psi}(0)}^{*}$*. In an abuse of notation, set* $\check{\Psi}(z)=\tilde{\Psi}(z)\tilde{\Psi}(0)^{-1}J$*, which is guaranteed to be a real matrix polynomial.*

*Now obtain a set of IRFs* $\check{\Theta}$ *and shock standard deviations* $\check{\sigma}$ *from* $\check{\Psi}(z)$*:*

*(a) For each* $j=1,\dots,n$*, if the* $(i_j,j)$ *element of* $\check{\Psi}(0)$ *is negative, flip the signs of all elements in the* $j$*-th column of* $\check{\Psi}(z)$*, and call the resulting matrix polynomial* $\check{\check{\Psi}}(z)$*. For each* $j=1,\dots,n$*, let* $\check{\sigma}_j$ *denote the* $(i_j,j)$ *element of* $\check{\check{\Psi}}(0)$*. Define* $\check{\sigma}=(\check{\sigma}_1,\dots,\check{\sigma}_n)$ *and* $\check{\Theta}(z)=\check{\check{\Psi}}(z)\,\mathrm{diag}(\check{\sigma})^{-1}$ *(if the inverse exists).*

*Then* $(\check{\Theta},\check{\sigma})\in\mathcal{S}(\Gamma)$*, provided that all elements of* $\check{\sigma}$ *are strictly positive.*

*On the other hand, if* $(\check{\Theta},\check{\sigma})\in\mathcal{S}(\Gamma)$ *is an arbitrary point in the identified set satisfying* $\det(\check{\Theta}(0))\neq 0$*, then* $(\check{\Theta},\check{\sigma})$ *can be obtained from* $(\Theta,\sigma)$ *as follows:*

1. *Start with the initial point* $(\Theta,\sigma)$ *and the associated polynomial* $\Psi(z)$ *defined above.*

2. *Apply an appropriate* finite *sequence of the above-mentioned transformations (i) or (ii), in an appropriate order, to* $\Psi(z)$*, resulting ultimately in a polynomial* $\check{\Psi}(z)$*.*

[49] If $\det(\Theta_0)=0$, some linear combination of $y_{1,t},\dots,y_{n,t}$ is perfectly predictable based on knowledge of shocks $\varepsilon_{t-1},\varepsilon_{t-2},\dots$ before time $t$. In most applications, this event ought to receive zero prior probability.

3. *Apply the above-mentioned operation (a) to* $\check{\Psi}(z)$*. The result is* $(\check{\Theta}, \check{\sigma})$*.*

REMARKS:

1. An initial point in the identified set can be obtained by following the procedure in Hannan (1970, pp. 64–66) and then applying transformation (a). This essentially corresponds to computing the Wold decomposition of $\{y_t\}$ and applying appropriate normalizations (Hannan, 1970, Thm. 2″, p. 158).

2. Transformation (ii) corresponds to "flipping the root" $\gamma_k$ of $\det(\Psi(z))$. If $\gamma_k$ is not real, transformation (ii) requires that we also flip the complex conjugate root $\overline{\gamma_k}$, since this ensures that the resulting matrix polynomial will be real after a rotation.

3. If the IRF parameter space $\Xi_\Theta$ were restricted to those IRFs that are invertible (cf. Section 2.3), then transformation (ii) would be unnecessary. In this case, the identified set for $\Psi(z) = \Theta(z)\,\mathrm{diag}(\sigma)$ can be obtained by taking any element in the set (e.g., the Wold IRFs) and applying all possible orthogonal rotations, i.e., transformation (i). This is akin to identification in SVARs, cf. Section 2.1 and Uhlig (2005, Prop. A.1).

4. The purpose of transformation (a) is to enforce the normalizations $\Theta_{i_j j,0} = 1$.

## A.3 Whittle likelihood and score

Let $V(\Psi)$ be an $nT \times nT$ symmetric block Toeplitz matrix consisting of $T \times T$ blocks of $n \times n$ matrices, where the $(s,t)$ block is given by $\sum_{\ell=0}^{q-(t-s)} \Psi_{\ell+(t-s)}\Psi_\ell'$ for $t \geq s$ and the sum is taken to equal 0 when $t > s + q$. Then the exact log likelihood function can be written

$$\log p_{Y|\Psi}(Y_T \mid \Psi) = -\tfrac{1}{2} nT \log(2\pi) - \tfrac{1}{2}\log\det(V(\Psi)) - \tfrac{1}{2} Y_T' V(\Psi)^{-1} Y_T. \tag{12}$$

This is what the Kalman filter described in the Online Appendix computes. For all $k = 0, 1, 2 \dots, T-1$, define the Fourier frequencies $\omega_k = 2\pi k/T$, the discrete Fourier transform (DFT) of the data $\tilde{y}_k = (2\pi T)^{-1/2} \sum_{t=1}^{T} e^{-\imath\omega_k(t-1)} y_t$, the DFT of the MA parameters $\tilde{\Psi}_k(\Psi) = \sum_{\ell=1}^{q+1} e^{-\imath\omega_k(\ell-1)} \Psi_{\ell-1}$, and the SVMA spectral density matrix $f_k(\Psi) = (2\pi)^{-1}\tilde{\Psi}_k(\Psi)\tilde{\Psi}_k(\Psi)^*$ at frequency $\omega_k$. The Whittle (1953) approximation to the log likelihood (12) is given by

$$\log p^W_{Y|\Psi}(Y_T \mid \Psi) = -nT\log(2\pi) - \frac{1}{2}\sum_{k=0}^{T-1}\left\{\log\det(f_k(\Psi)) + \tilde{y}_k^*[f_k(\Psi)]^{-1}\tilde{y}_k\right\}.$$

The approximation is obtained by substituting $V(\Psi) \approx 2\pi\Delta F(\Psi)\Delta^*$ in (12) (Brockwell & Davis, 1991, Prop. 4.5.2). Here $\Delta$ is an $nT \times nT$ unitary matrix with $(s,t)$ block equal to $T^{-1/2}e^{\imath\omega_{s-1}(t-1)}I_n$. $F(\Psi)$ is a block diagonal $nT \times nT$ matrix with $(s,s)$ block equal to $f_s(\Psi)$. The Whittle log likelihood is computationally cheap because $\{\tilde{y}_k, \tilde{\Psi}_k(\Psi)\}_{0\leq k\leq T-1}$ can be computed efficiently using the Fast Fourier Transform (Hansen & Sargent, 1981, Sec. 2b; Brockwell & Davis, 1991, Ch. 10.3).[50]

Now I derive the gradient of the Whittle log likelihood. For all $k = 0, 1, \ldots, T-1$, define $C_k(\Psi) = [f_k(\Psi)]^{-1} - [f_k(\Psi)]^{-1}\tilde{y}_k\tilde{y}_k^*[f_k(\Psi)]^{-1}$ and its Discrete Fourier Transform $\tilde{C}_k(\Psi) = \sum_{\ell=1}^{T} e^{-\imath\omega_k(\ell-1)}C_{\ell-1}(\Psi)$.[51] Finally, let $\tilde{C}_k(\Psi) = \tilde{C}_{T+k}(\Psi)$ for $k = -1, -2, \ldots, 1-T$.

**Lemma 2.**

$$\frac{\log p^W_{Y|\Psi}(Y_T \mid \Psi)}{\partial \Psi_\ell} = -\sum_{\tilde{\ell}=0}^{q} \mathrm{Re}[\tilde{C}_{\tilde{\ell}-\ell}(\Psi)]\Psi_{\tilde{\ell}}, \quad \ell = 0, 1, \ldots, q. \tag{13}$$

The lemma gives the score with respect to $\Psi$. Since $\Psi_\ell = \Theta_\ell \operatorname{diag}(\sigma)$, the chain rule gives the score with respect to $\Theta$ and $\log\sigma$.

## A.4 Asymptotic theory: Mathematical details

### A.4.1 Whittle likelihood for a $q$-dependent process

I now define the Whittle ACF likelihood for $q$-dependent processes mentioned in Section 5.2. Consider the spectral density for a $q$-dependent process parametrized in terms of its ACF:

$$f(\omega;\Gamma) = \frac{1}{2\pi}\left(\Gamma(0) + \sum_{k=1}^{q}\{e^{-\imath k\omega}\Gamma(k) + e^{\imath k\omega}\Gamma(k)'\}\right), \quad \Gamma \in \mathbb{T}_{n,q},$$

$$\mathbb{T}_{n,q} = \Bigg\{\{\Gamma(k)\}_{0\leq k\leq q} \colon \Gamma(\cdot) \in \mathbb{R}^{n\times n},\ \Gamma(0) = \Gamma(0)', \\ \inf_{\omega\in[0,\pi)} \det\left(\Gamma(0) + \sum_{k=1}^{q}\{e^{-\imath\omega k}\Gamma(k) + e^{\imath\omega k}\Gamma(k)'\}\right) > 0\Bigg\}.$$

[50] As noted by Hansen & Sargent (1981, p. 32), the computation time can be halved by exploiting $\tilde{y}_{T-k} = \overline{\tilde{y}_k}$ and $f_{T-k}(\Psi) = \overline{f_k(\Psi)}$ for $k = 1, 2, \ldots, T$.

[51] Again, computation time can be saved by exploiting $C_{T-k}(\Psi) = \overline{C_k(\Psi)}$ for $k = 1, 2, \ldots, T$.

Let $\hat{\Gamma}(k) = T^{-1}\sum_{t=1}^{T-k} y_{t+k}y_t'$, $k = 0, 1, \dots, T-1$, be the $k$-th sample autocovariance, and set $\hat{\Gamma}(k) = \hat{\Gamma}(-k)'$ for $k = -1, -2, \dots, 1-T$. Define the periodogram

$$\hat{I}(\omega) = \frac{1}{2\pi T}\left(\sum_{t=1}^{T} e^{-\imath t\omega} y_t\right)\left(\sum_{t=1}^{T} e^{\imath t\omega} y_t'\right) = \frac{1}{2\pi}\sum_{k=-(T-1)}^{T-1} e^{-\imath k\omega}\hat{\Gamma}(k), \quad \omega \in [-\pi, \pi].$$

The Whittle ACF log likelihood is (up to a constant) given by

$$\log p^W_{Y|\Gamma}(Y_T \mid \Gamma) = -\frac{T}{4\pi}\int_{-\pi}^{\pi} \log\det(f(\omega;\Gamma))\, d\omega - \frac{T}{4\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{f(\omega;\Gamma)^{-1}\hat{I}(\omega)\}\, d\omega.$$

As in Appendix A.3, it is common to use a discretized Whittle log likelihood that replaces integrals with corresponding discretized sums. The proof of Proposition 1 shows that posterior consistency also holds when the discretized Whittle likelihood is used.

### A.4.2 Posterior consistency for Wold parameters

The proof of Proposition 1 relies on the following posterior consistency result for the Wold IRFs and prediction covariance matrix. The result in this subsection concerns (invertible) reduced-form IRFs, not (possibly noninvertible) structural IRFs.

Fix a finite $q \in \mathbb{N}$, and let $\beta_0(L) = I_n + \sum_{\ell=1}^{q}\beta_{0,\ell}L^\ell$ and $\Sigma_0$ denote the MA lag polynomial and prediction covariance matrix, respectively, in the Wold decomposition (Hannan, 1970, Thm. 2″, p. 158) of a $q$-dependent stationary $n$-dimensional process with ACF (out to lag $q$) given by $\{\Gamma_0(k)\}_{0\le k\le q} \in \mathbb{T}_{n,q}$. That is, $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,q}) \in \mathbb{B}_{n,q}$ and $\Sigma_0 \in \mathbb{S}_n$ are the unique parameters such that $\Gamma_0(k) = \sum_{\ell=0}^{q-k}\beta_{0,\ell+k}\Sigma_0\beta_{0,\ell}'$ for $0 \le k \le q$, defining $\beta_{0,0} = I_n$. Here $\mathbb{S}_n$ denotes the space of symmetric positive definite $n \times n$ matrices, while

$$\mathbb{B}_{n,q} = \left\{\beta = (\beta_1, \dots, \beta_q) \in \mathbb{R}^{n\times nq}\colon\ \det(\Phi(z;\beta)) \neq 0\ \forall\, |z| \le 1\right\},$$

$$\Phi(z;\beta) = I_n + \sum_{\ell=1}^{q}\beta_\ell z^\ell, \quad z \in \mathbb{C}.$$

Define the MA spectral density parametrized in terms of $(\beta, \Sigma)$:

$$\tilde{f}(\omega;\beta,\Sigma) = \frac{1}{2\pi}\Phi(e^{-\imath\omega};\beta)\Sigma\Phi(e^{\imath\omega};\beta)', \quad \omega \in [-\pi,\pi], \quad (\beta,\Sigma) \in \mathbb{B}_{n,q}\times\mathbb{S}_n.$$

Using the notation in Appendix A.4.1 for the periodogram $\hat{I}(\omega)$, the Whittle MA log likelihood is (up to a constant) given by

$$\log p^W_{Y|\beta,\Sigma}(Y_T \mid \beta, \Sigma) = -\frac{T}{4\pi}\int_{-\pi}^{\pi} \log\det(\tilde{f}(\omega;\beta,\Sigma))\,d\omega - \frac{T}{4\pi}\int_{-\pi}^{\pi} \operatorname{tr}\{\tilde{f}(\omega;\beta,\Sigma)^{-1}\hat{I}(\omega)\}\,d\omega.$$

The result below also holds for the discretized Whittle likelihood (cf. Appendix A.4.1).

I now state the posterior consistency result for $(\beta_0, \Sigma_0)$. Let $\Pi_{\beta,\Sigma}(\cdot)$ be a prior measure for $(\beta_0, \Sigma_0)$ on $\mathbb{B}_{n,q} \times \mathbb{S}_n$. Define the Whittle posterior measure

$$P^W_{\beta,\Sigma|Y}(\mathcal{A} \mid Y_T) = \frac{\int_{\mathcal{A}} p^W_{Y|\beta,\Sigma}(Y_T \mid \beta,\Sigma)\Pi_{\beta,\Sigma}(d(\beta,\Sigma))}{\int_{\mathbb{B}_{n,q}\times\mathbb{S}_n} p^W_{Y|\beta,\Sigma}(Y_T \mid \beta,\Sigma)\Pi_{\beta,\Sigma}(d(\beta,\Sigma))}$$

for any measurable set $\mathcal{A} \subset \mathbb{B}_{n,q} \times \mathbb{S}_n$. Note that the lemma below does not require the true data distribution to be Gaussian or $q$-dependent.

**Lemma 3.** *Let Assumption 3 hold. Assume that the pseudo-true parameters $(\beta_0, \Sigma_0) \in \mathbb{B}_{n,q} \times \mathbb{S}_n$ are in the support of the prior $\Pi_{\beta,\Sigma}(\cdot)$. Then, for any neighborhood $\tilde{\mathcal{U}}$ of $(\beta_0, \Sigma_0)$ in $\mathbb{B}_{n,q} \times \mathbb{S}_n$, we have $P^W_{\beta,\Sigma|Y}(\tilde{\mathcal{U}} \mid Y_T) \xrightarrow{p} 1$ under the true probability measure of the data.*

# B Proofs

## B.1 Proof of Lemma 1

By the triangle inequality,

$$\|P_{\theta|Y}(\cdot \mid Y_T) - \Pi_{\theta|\Gamma}(\cdot \mid \hat{\Gamma})\|_{L_1} \leq \|\Pi_{\theta|\Gamma}(\cdot \mid \hat{\Gamma}) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1} + \|P_{\theta|Y}(\cdot \mid Y_T) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1}.$$

If $\hat{\Gamma} \xrightarrow{p} \Gamma_0$, the first term above tends to 0 in probability by assumption (i) and the continuous mapping theorem. Hence, the statement of the lemma follows if I can show that the second term above tends to 0 in probability.

Let $\epsilon > 0$ be arbitrary. By assumption (i), there exists a neighborhood $\mathcal{U}$ of $\Gamma_0$ in $\Xi_\Gamma$ such that $\|\Pi_{\theta|\Gamma}(\cdot \mid \Gamma) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1} < \epsilon/2$ for all $\Gamma \in \mathcal{U}$. By assumption (ii), $P_{\Gamma|Y}(\mathcal{U}^c \mid Y_T) < \epsilon/4$ w.p.a. 1. The decomposition (9) then implies

$$\begin{aligned}\|P_{\theta|Y}(\cdot \mid Y_T) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1} &= \left\| \int [\Pi_{\theta|\Gamma}(\cdot \mid \Gamma) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)] P_{\Gamma|Y}(d\Gamma \mid Y_T) \right\|_{L_1} \\ &\leq \int_{\mathcal{U}} \|\Pi_{\theta|\Gamma}(\cdot \mid \Gamma) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1} P_{\Gamma|Y}(d\Gamma \mid Y_T)\end{aligned}$$

$$
\begin{aligned}
&+\int_{\mathcal{U}^c} \|\Pi_{\theta|\Gamma}(\cdot \mid \Gamma) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0)\|_{L_1} P_{\Gamma|Y}(d\Gamma \mid Y_T) \\
&\leq \int_{\mathcal{U}} \frac{\epsilon}{2} P_{\Gamma|Y}(d\Gamma \mid Y_T) + 2P_{\Gamma|Y}(\mathcal{U}^c \mid Y_T) \\
&\leq \frac{\epsilon}{2} + 2\frac{\epsilon}{4}
\end{aligned}
$$

w.p.a. 1. Here I use that the $L_1$ distance between probability measures is bounded by 2. □

## B.2 Proof of Proposition 1

The proof exploits the one-to-one mapping between the ACF $\Gamma_0$ and the Wold parameters $(\beta_0, \Sigma_0)$ defined in Appendix A.4.2, which allows me to use Lemma 3 to infer posterior consistency for $\Gamma_0$ under the Whittle likelihood.

Let $M \colon \mathbb{T}_{n,q} \to \mathbb{B}_{n,q} \times \mathbb{S}_n$ denote the function that maps a $q$-dependent ACF $\Gamma(\cdot)$ into its Wold representation $(\beta(\Gamma), \Sigma(\Gamma))$ (Hannan, 1970, Thm. 2″, p. 158). By construction, the map $M(\cdot)$ is continuous (and measurable). The inverse map $M^{-1}(\cdot)$ is given by $\Gamma(k) = \sum_{\ell=0}^{q-k} \beta_{\ell+k} \Sigma \beta_\ell'$ (with $\beta_0 = I_n$) and so also continuous. The prior $\Pi_\Gamma(\cdot)$ for the ACF $\Gamma$ induces a particular prior measure for the Wold parameters $(\beta, \Sigma)$ on $\mathbb{B}_{n,q} \times \mathbb{S}_n$ given by $\Pi_{\beta,\Sigma}(\mathcal{A}) = \Pi_\Gamma(M^{-1}(\mathcal{A}))$ for any measurable set $\mathcal{A}$. Let $P^W_{\beta,\Sigma|Y}(\cdot \mid Y_T)$ be the posterior measure for $(\beta, \Sigma)$ computed using the induced prior $\Pi_{\beta,\Sigma}(\cdot)$ and the Whittle MA likelihood $p^W_{Y|\beta,\Sigma}(Y_T \mid \beta, \Sigma)$, cf. Appendix A.4.2.

I first show that the induced posterior for $(\beta_0, \Sigma_0)$ is consistent. Let $\tilde{\mathcal{U}}$ be any neighborhood of $(\beta_0, \Sigma_0) = M(\{\Gamma_0(k)\}_{0 \leq k \leq q})$ in $\mathbb{B}_{n,q} \times \mathbb{S}_n$. Since $M(\cdot)$ is continuous, $M^{-1}(\tilde{\mathcal{U}})$ is a neighborhood of $\{\Gamma_0(k)\}_{0 \leq k \leq q}$ in $\mathbb{T}_{n,q}$. Hence, since $\{\Gamma_0(k)\}_{0 \leq k \leq q}$ is in the support of $\Pi_\Gamma(\cdot)$, $(\beta_0, \Sigma_0)$ is in the support of $\Pi_{\beta,\Sigma}(\cdot)$: $\Pi_{\beta,\Sigma}(\tilde{\mathcal{U}}) = \Pi_\Gamma(M^{-1}(\tilde{\mathcal{U}})) > 0$. Due to Assumption 3 and the fact that $(\beta_0, \Sigma_0)$ is in the support of $\Pi_{\beta,\Sigma}(\cdot)$, Lemma 3 implies that $P_{\beta,\Sigma|Y}(\tilde{\mathcal{U}} \mid Y_T) \xrightarrow{p} 1$ for any neighborhood $\tilde{\mathcal{U}}$ of $(\beta_0, \Sigma_0)$ in $\mathbb{B}_{n,q} \times \mathbb{S}_n$.

I now prove posterior consistency for $\Gamma_0$. Since $\tilde{f}(\omega; M(\Gamma)) = f(\omega; \Gamma)$ for all $\omega \in [-\pi, \pi]$ and $\Gamma \in \mathbb{T}_{n,q}$, we have $p^W_{Y|\beta,\Sigma}(Y_T \mid M(\Gamma)) = p^W_{Y|\Gamma}(Y_T \mid \Gamma)$ for all $\Gamma \in \mathbb{T}_{n,q}$. Consequently, $P^W_{\Gamma|Y}(\mathcal{A} \mid Y_T) = P^W_{\beta,\Sigma|Y}(M(\mathcal{A}) \mid Y_T)$ for all measurable sets $\mathcal{A}$. Let $\mathcal{U}$ be an arbitrary neighborhood of $\{\Gamma_0(k)\}_{0 \leq k \leq q}$ in $\mathbb{T}_{n,q}$. Since $M^{-1}(\cdot)$ is continuous at $(\beta_0, \Sigma_0)$, the set $\tilde{\mathcal{U}} = M(\mathcal{U})$ is a neighborhood of $(\beta_0, \Sigma_0)$ in $\mathbb{B}_{n,q} \times \mathbb{S}_n$. It follows that

$$
P^W_{\Gamma|Y}(\mathcal{U} \mid Y_T) = P^W_{\beta,\Sigma|Y}(\tilde{\mathcal{U}} \mid Y_T) \xrightarrow{p} 1.
$$

The proof of Lemma 3 shows that the discretized Whittle posterior is also consistent. □

## B.3 Proof of Proposition 2

By Lemma 1, I just need to verify posterior consistency. The calculation in Eqn. 11 of Moon & Schorfheide (2012) shows that the Whittle posterior $P^W_{\theta|Y}(\cdot \mid Y_T)$ satisfies the decomposition (9), where the posterior measure for $\Gamma$ is given by $P^W_{\Gamma|Y}(\cdot \mid Y_T)$. By Proposition 1, the latter posterior is consistent provided that $\Gamma_0$ is in the support of the induced prior $\Pi_\Gamma(\cdot)$.

$\Gamma_0$ is indeed in the support of $\Pi_\Gamma(\cdot)$, for the following reason. Let $\Gamma(\Theta, \sigma)$ denote the map (6) from structural parameters $(\Theta, \sigma) \in \Xi_\Theta \times \Xi_\sigma$ to ACFs $\Gamma \in \mathbb{T}_{n,q}$. There exists a (non-unique) set of IRFs and shock standard deviations $(\check{\Theta}, \check{\sigma}) \in \Xi_\Theta \times X_\sigma$ such that $\Gamma_0 = \Gamma(\check{\Theta}, \check{\Sigma})$ (Hannan, 1970, pp. 64–66). Let $\mathcal{U}$ be an arbitrary neighborhood of $\Gamma_0$ in $\mathbb{T}_{n,q}$. The map $\Gamma(\cdot, \cdot)$ is continuous, so $\Gamma^{-1}(\mathcal{U})$ is a neighborhood of $(\check{\Theta}, \check{\sigma})$ in $\Xi_\Theta \times \Xi_\sigma$. Because $\Pi_{\Theta,\sigma}(\cdot)$ has full support on $\Xi_\Theta \times \Xi_\sigma$, we have $\Pi_\Gamma(\mathcal{U}) = \Pi_{\Theta,\sigma}(\Gamma^{-1}(\mathcal{U})) > 0$. Since the neighborhood $\mathcal{U}$ was arbitrary, $\Gamma_0$ lies in the support of the induced prior $\Pi_\Gamma(\cdot)$. □

## B.4 Proofs of Lemmas 2 and 3 and Proposition 3

Please see the Online Appendix.

# References


Adjemian, S., Bastani, H., Juillard, M., Karamé, F., Mihoubi, F., Perendia, G., Pfeifer, J., Ratto, M. & Villemot, S. (2011). Dynare: Reference Manual, Version 4. Dynare Working Papers, 1, CEPREMAP.

Alessi, L., Barigozzi, M. & Capasso, M. (2011). Non-Fundamentalness in Structural Econometric Models: A Review. *International Statistical Review* 79(1), 16–47.

Arias, J. E., Rubio-Ramírez, J. F. & Waggoner, D. F. (2016). Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications. Working paper.

Barnichon, R. & Matthes, C. (2016). Gaussian Mixture Approximations of Impulse Responses and The Non-Linear Effects of Monetary Shocks. Working paper.

Baumeister, C. & Hamilton, J. D. (2015a). Optimal Inference about Impulse-Response Functions and Historical Decompositions in Incompletely Identified Structural Vector Autoregressions. Working paper.

Baumeister, C. & Hamilton, J. D. (2015b). Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information. *Econometrica* 83(5), 1963–1999.

Baumeister, C. & Hamilton, J. D. (2015c). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. Working paper.

Beaudry, P., Fève, P., Guay, A. & Portier, F. (2015). When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks. National Bureau of Economic Research Working Paper No. 21466.

Beaudry, P. & Portier, F. (2014). News-Driven Business Cycles: Insights and Challenges. *Journal of Economic Literature* 52(4), 993–1074.

Blanchard, O. J., L'Huillier, J. P. & Lorenzoni, G. (2013). News, Noise, and Fluctuations: An Empirical Exploration. *American Economic Review* 103(7), 3045–3070.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.). Springer Series in Statistics. Springer.

Chan, J. C. C., Eisenstat, E. & Koop, G. (2015). Large Bayesian VARMAs. Working paper.

Chib, S. & Greenberg, E. (1994). Bayes inference in regression models with ARMA $(p, q)$ errors. *Journal of Econometrics* 64, 183–206.

Christiano, L. J. & Vigfusson, R. J. (2003). Maximum likelihood in the frequency domain: the importance of time-to-plan. *Journal of Monetary Economics* 50(4), 789–815.

Del Negro, M. & Schorfheide, F. (2004). Priors from General Equilibrium Models for VARs. *International Economic Review* 45(2), 643–673.

Drèze, J. H. & Richard, J.-F. (1983). Bayesian Analysis of Simultaneous Equation Models. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of Econometrics*, Vol. 1, Ch. 9, 517–598. North-Holland.

Dwyer, M. (1998). Impulse Response Priors for Discriminating Structural Vector Autoregressions. Working paper.

Fernald, J. (2014). A Quarterly, Utilization-Adjusted Series on Total Factor Productivity. Federal Reserve Bank of San Francisco Working Paper No. 2012-19.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J. & Watson, M. W. (2007). ABCs (and Ds) of Understanding VARs. *American Economic Review* 97(3), 1021–1026.

Fève, P. & Jidoud, A. (2012). Identifying News Shocks from SVARs. *Journal of Macroeconomics* 34(4), 919–932.

Forni, M. & Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics* 66, 124–136.

Forni, M., Gambetti, L., Lippi, M. & Sala, L. (2013). Noisy News in Business Cycles. Working paper.

Forni, M., Gambetti, L. & Sala, L. (2014). No News in Business Cycles. *Economic Journal* 124(581), 1168–1191.

Forni, M., Giannone, D., Lippi, M. & Reichlin, L. (2009). Opening the Black Box: Structural Factor Models With Large Cross Sections. *Econometric Theory* 25(05), 1319–1347.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. & Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman & Hall/CRC Texts in Statistical Science. CRC Press.

Giacomini, R. & Kitagawa, T. (2015). Robust inference about partially identified SVARs. Working paper.

Giannone, D. & Reichlin, L. (2006). Does Information Help Recovering Structural Shocks from Past Observations? *Journal of the European Economic Association* 4(2–3), 455–465.

Gordon, S. & Boccanfuso, D. (2001). Learning from Structural Vector Autoregression Models. Working paper.

Gospodinov, N. & Ng, S. (2015). Minimum Distance Estimation of Possibly Noninvertible Moving Average Models. *Journal of Business & Economic Statistics* 33(3), 403–417.

Gustafson, P. (2015). *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

Hannan, E. (1970). *Multiple Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Hansen, L. P. & Sargent, T. (1981). Exact linear rational expectations models: Specification and estimation. Federal Reserve Bank of Minneapolis Staff Report No. 71.

Hansen, L. P. & Sargent, T. J. (1991). Two Difficulties in Interpreting Vector Autoregressions. In L. P. Hansen & T. J. Sargent (Eds.), *Rational Expectations Econometrics*, Underground Classics in Economics, Ch. 4, 77–119. Westview Press.

Hoffman, M. D. & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15, 1593–1623.

Ingram, B. F. & Whiteman, C. H. (1994). Supplanting the 'Minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors. *Journal of Monetary Economics* 34(3), 497–510.

Ito, T. & Quah, D. (1989). Hypothesis Testing with Restricted Spectral Density Matrices, with an Application to Uncovered Interest Parity. *International Economic Review* 30(1), 203–215.

Jordà, O. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review* 95(1), 161–182.

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

Klaeffing, M. (2003). Monetary Policy Shocks – A Nonfundamental Look at the Data. European

Central Bank Working Paper No. 228.

Kocięcki, A. (2010). A Prior for Impulse Responses in Bayesian Structural VAR Models. *Journal of Business & Economic Statistics* 28(1), 115–127.

Komunjer, I. & Ng, S. (2011). Dynamic Identification of Dynamic Stochastic General Equilibrium Models. *Econometrica* 79(6), 1995–2032.

Lanne, M. & Saikkonen, P. (2013). Noncausal Vector Autoregression. *Econometric Theory* 29(03), 447–481.

Leeper, E. M., Walker, T. B. & Yang, S.-C. S. (2013). Fiscal Foresight and Information Flows. *Econometrica* 81(3), 1115–1145.

Lippi, M. & Reichlin, L. (1994). VAR analysis, nonfundamental representations, Blaschke matrices. *Journal of Econometrics* 63(1), 307–325.

Lütkepohl, H. (2014). Fundamental Problems with Nonfundamental Shocks. In N. Haldrup, M. Meitz, & P. Saikkonen (Eds.), *Essays in Nonlinear Time Series Econometrics*, Ch. 8, 198–214. Oxford University Press.

Mertens, K. & Ravn, M. O. (2010). Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach. *The Economic Journal* 120(544), 393–413.

Mertens, K. & Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review* 103(4), 1212–1247.

Moon, H. R. & Schorfheide, F. (2012). Bayesian and Frequentist Inference in Partially Identified Models. *Econometrica* 80(2), 755–782.

Müller, U. K. (2013). Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica* 81(5), 1805–1849.

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, Handbooks of Modern Statistical Methods, Ch. 5, 113–162. CRC Press.

Phillips, P. (1989). Partially Identified Econometric Models. *Econometric Theory* 5(02), 181–240.

Poirier, D. J. (1998). Revising Beliefs in Nonidentified Models. *Econometric Theory* 14(04), 483–509.

Qu, Z. & Tkachenko, D. (2012a). Frequency Domain Analysis of Medium Scale DSGE Models with Application to Smets and Wouters (2007). In N. Balke, F. Canova, F. Milani, & M. A. Wynne (Eds.), *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*, Vol. 28 of *Advances in Econometrics*, 319–385. Emerald Group Publishing.

Qu, Z. & Tkachenko, D. (2012b). Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models. *Quantitative Economics* 3(1), 95–132.

Ramamoorthi, R. V., Sriram, K. & Martin, R. (2015). On Posterior Concentration in Misspecified Models. *Bayesian Analysis* 10(4), 759–789.

Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation. Draft of chapter to appear in *Handbook of Macroeconomics*, Vol. 2.

Rotemberg, J. & Woodford, M. (1997). An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy. In B. S. Bernanke & J. Rotemberg (Eds.), *NBER Macroeconomics Annual*, Vol. 12, 297–346. MIT Press.

Sala, L. (2015). DSGE Models in the Frequency Domain. *Journal of Applied Econometrics* 30(2), 219–240.

Shalizi, C. R. (2009). Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics* 3, 1039–1074.

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

Sims, C. A. & Zha, T. (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review* 39(4), 949–968.

Sims, C. A. & Zha, T. (2006). Does Monetary Policy Generate Recessions? *Macroeconomic Dynamics* 10(02), 231–272.

Sims, E. R. (2012). News, Non-Invertibility, and Structural VARs. In N. Balke, F. Canova, F. Milani, & M. A. Wynne (Eds.), *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*, Vol. 28 of *Advances in Econometrics*, 81–135. Emerald Group Publishing.

Stock, J. H. & Watson, M. W. (2008). Recent Developments in Structural VAR Modeling. NBER Summer Institute Minicourse 2008, "What's New in Econometrics – Time Series", Lecture 7.

Stock, J. H. & Watson, M. W. (2012). Disentangling the Channels of the 2007–09 Recession. *Brookings Papers on Economic Activity* (Spring), 81–135.

Stock, J. H. & Watson, M. W. (2016). Factor Models and Structural Vector Autoregressions in Macroeconomics. Draft of chapter to appear in *Handbook of Macroeconomics*, Vol. 2.

Tamaki, K. (2008). The Bernstein-von Mises Theorem for Stationary Processes. *Journal of the Japan Statistical Society* 38(2), 311–323.

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.

Watson, M. W. (1994). Vector autoregressions and cointegration. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of Econometrics*, Vol. 4, Ch. 47, 2843–2915. North-Holland.

Whittle, P. (1953). The Analysis of Multiple Stationary Time Series. *Journal of the Royal Statistical Society, Series B (Methodological)* 15(1), 125–139.