Bayesian Inference on Structural Impulse Response Functions

Mikkel Plagborg-Møller∗

May 10, 2016

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. To rapidly simulate from the posterior of the impulse responses, I develop 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. I demonstrate the usefulness of my method in a simulation study and in an empirical application that estimates the effects of technological news shocks on the U.S. business cycle.

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

∗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, José Luis Montiel Olea, Frank Schorfheide, Elie Tamer, Harald Uhlig, and seminar participants at Cambridge, Chicago Booth, Columbia, Harvard Economics and Kennedy School, MIT, 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 Theorem 3. I am indebted to Gary Chamberlain, Gita Gopinath, Anna Mikusheva, Neil Shephard, and Jim Stock for their help and guidance.
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 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. My 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 possible, I propose a multivariate Gaussian prior that facilitates graphical prior elicitation: Sketch the prior means for each impulse response in a plot, then place prior confidence bands around the means, and finally specify prior information about the smoothness (i.e., prior correlation) of the IRFs. In particular, researchers can easily and 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, noninvertible 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*.

I demonstrate the practical usefulness of the SVMA method through a simulation exercise and an empirical application. The simulations show that prior information about the smoothness of IRFs can sharpen posterior inference about unknown features of the IRFs, since smoother IRFs have fewer effective free parameters. Prior information about smoothness has not been explicitly exploited in the SVAR literature, because the shapes and smoothness of SVAR IRFs are complicated functions of the underlying SVAR parameters.

My empirical application estimates the effects of technological news shocks on the U.S. business cycle, using data on productivity, output, and the real interest rate. 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 noninvertible IRFs without dogmatically imposing a particular Dynamic Stochastic General Equilibrium (DSGE) model. I use the sticky-price DSGE model in *E. Sims (2012)* to guide prior elicitation. My 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 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. I overcome the computational challenges of the SVMA model by simulating from the posterior using Hamiltonian Monte Carlo (HMC), a Markov Chain Monte Carlo method that is well-suited to high-dimensional models (*Neal, 2011*). 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 information 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. I show that, asymptotically, the role of the data is to pin down the true autocovariances of the data, which in turn pins down the reduced-form (Wold) impulse responses; all other information about structural impulse responses comes from the prior. Furthermore, I prove that the approximation error incurred by using the Whittle likelihood is negligible asymptotically.

As a key step in the asymptotic analysis, I show that the posterior distribution for the autocovariance function of essentially any covariance stationary time series is consistent for the true value. While the posterior is computed under the working assumption that the data is Gaussian and $q$-dependent, consistency obtains under general misspecification of the Whittle likelihood. Existing time series results on posterior consistency assume well-specified likelihood functions. The only assumptions I place on the data generating process are nonparametric covariance stationarity and weak dependence conditions, and the prior is unrestricted except its support must contain the true autocovariance function.

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. The 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, by expanding the vector of observed time series.

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

¹ [http://scholar.harvard.edu/plagborg/irf_bayes](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.
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 (2015) 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 impose parametric structures on the IRFs, while I show how to maintain computational tractability with potentially unrestricted IRFs.

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 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 on IRFs. 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 illustrates the SVMA approach through a small simulation study. Section 5 empirically estimates the role of technological news shocks in the U.S. business cycle. Section 6 derives the large-sample limit of the posterior distribution for a large class of partially identified models that includes the SVMA model. Section 7 shows that popular SVAR restrictions can be imposed in the SVMA framework. Section 8 suggests topics for future research. Applied readers may want to focus on Sections 2 to 5. Technical details are relegated to Appendix A; in particular, notation is defined in Appendix A.1. Proofs can be found in Appendix B.

\(^3\)Standard errors in Hansen & Sargent (1981) are only valid when the prior restrictions point identify the IRFs. Barnichon & Matthes (2015) 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.
2 Model, invertibility, and prior elicitation

In this section I describe the SVMA model and my method for imposing priors on IRFs. After reviewing the SVAR model and its shortcomings, I discuss the SVMA model, whose parameters are the IRFs of observed variables to unobserved shocks. Because the SVMA model does not restrict the IRFs to be invertible a priori, it can be applied to more empirical settings than the SVAR approach. The IRFs are under-identified, as they are in SVAR analysis. The lack of identification necessitates the use of prior information, which I impose by placing a prior distribution on the IRFs that lets researchers flexibly and transparently exploit all types of prior information about IRFs.

2.1 SVARs and their shortcomings

I begin with a brief review of Structural Vector Autoregression (SVAR) estimation of impulse response functions. 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 is based on Frisch’s (1933) impulse-propagation paradigm, which 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. It has long been recognized that – in a linear setting – this paradigm is well captured by the Structural Vector Moving Average (SVMA) model (Hansen & Sargent, 1981; Watson, 1994, Sec. 4)

\[ y_t = \Theta(L)\varepsilon_t, \quad \Theta(L) = \sum_{\ell=0}^{\infty} \Theta_\ell L^\ell, \]  

(1)

where \( L \) denotes the lag operator, \( y_t = (y_{1,t}, \ldots, y_{n,t})' \) is an \( n \)-dimensional vector of observed macro variables, and the structural shocks \( \varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})' \) form a martingale difference sequence with \( E(\varepsilon_t\varepsilon'_t) = \text{diag}(\sigma)^2, \quad \sigma = (\sigma_1, \ldots, \sigma_n)' \). Most linearized discrete-time macro models can be written in SVMA form. \( \Theta_{i,j,\ell} \), 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 \((\Theta_{i,j,\ell})_{\ell\geq 0}\) is thus a key object of scientific interest in macroeconomics (Ramey, 2015).

For computational reasons, most researchers follow Sims (1980) and estimate structural
IRFs and shocks 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 \( A_1, \ldots, A_m \) 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. Denote the reduced-form (Wold) forecast error by \( u_{t|t-1} = y_t - \text{proj}(y_t \mid y_{t-1}, y_{t-2}, \ldots) = H \varepsilon_t \), where “proj” denotes population linear projection. Then the only information available from second moments of the data to identify \( H \) and \( \sigma \) is that

\[ E(u_{t|t-1} u_{t|t-1}' | t-1) = H \text{diag}(\sigma)^2 H' \tag{5} \]

\(^5\) 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 \( \Theta_{ij,\ell} \) at long horizons \( \ell \) are extrapolated from the short-run correlations of the data. Hence, the overall 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 of the IRFs, such SVAR identification schemes are known to impose additional unintended and unacknowledged prior

---

\(^4\) 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).

\(^5\) Equivalently, if \( E(u_{t|t-1} u_{t|t-1}' | t-1) = J J' \) is 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).

\(^6\) 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 \( A_1, \ldots, A_m \). See Geweke (1988, Sec. 2) for an illustration in the univariate case.

\(^7\) Ramey (2015), Stock & Watson (2015), and Uhlig (2015) review SVAR identification schemes.
information about IRFs.\footnote{See Arias, Rubio-Ramírez & Waggoner (2014, Sec. 5) and Baumeister & Hamilton (2015b, Sec. 3). For a trivial example, consider the AR(1) model $y_t = A_1 y_{t-1} + \varepsilon_t$ with $n = 1$ and $|A_1| < 1$. The IRF is $\Theta_\ell = A_1^\ell$, so imposing the sign restriction $\Theta_1 \geq 0$ implicitly also restricts $\Theta_\ell \geq 0$ for all $\ell \geq 2$.}

A second drawback of the SVAR model is the invertibility problem. The defining property of the SVAR model (2) is that the structural shocks $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})'$ can be recovered linearly from the history $(y_t, y_{t-1}, \ldots)$ of observed data, given knowledge of $H$ and $\sigma$. This invertibility assumption – that the time-$t$ shocks can be recovered from current and past, but not future, values of the observed data – is arbitrary and may be violated if the econometrician does not observe all variables relevant to the decisions of forward-looking economic agents. Indeed, the literature has demonstrated that data generated by interesting macroeconomic models, including models with news or noise shocks, cannot be represented as a SVAR for this reason. Section 2.3 discusses invertibility in greater detail and provides references.

To overcome the drawbacks of the SVAR model, I return to the basics and infer IRFs directly from the SVMA representation (1) of the data. The SVMA parameters are underidentified, so prior information must be imposed to learn about unknown features of the IRFs. Conveniently, the parameters of the SVMA model are the IRFs themselves, so all types of prior information about IRFs can be imposed easily and transparently. Moreover, because the IRFs $\Theta(L)$ are unrestricted, the structural shocks $\varepsilon_t$ need not be recoverable from only current and past values of the data. Hence, the SVMA model can handle applications in which the data may not have a SVAR representation, as in the examples described above.

### 2.2 SVMA model

I now discuss the SVMA model assumptions 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 $y_t = (y_{1,t}, \ldots, y_{n,t})'$ are driven by current and lagged values of unobserved, unpredictable shocks $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})'$ (Hansen & Sargent, 1981). Although the shocks are unobserved, the researcher must have some degree of prior knowledge about their nature in order to estimate the SVMA parameters, as described in Section 2.5. 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, cf. Section 8.
**Assumption 1** (SVMA model).

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

where \( L \) is the lag operator, \( q \) is the finite MA lag length, and \( \Theta_0, \Theta_1, \ldots, \Theta_q \) 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 \( E(\varepsilon^2_{j,t}) = \sigma^2_j, \) 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 or Box-Jenkins methods. To fit persistent data \( q \) must be relatively large, which my computational strategy in Section 3 is well-suited for. The assumption that \( y_t \) has mean zero is made for notational convenience and can easily be relaxed. Unlike in reduced-form Vector Autoregressive Moving Average (VARMA) modeling, the SVMA model allows \( \Theta_0 \neq I_n. \)

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 structure (3) is compatible with an underlying SVAR model \( D(L) y_t = \varepsilon_t \) (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, as explained in Section 2.3. Assumption 1 imposes stationary, linear dynamics with time-invariant parameters, which is restrictive but standard in the SVAR literature.\(^9\) The condition that \( \varepsilon_t \) form 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.\(^{10}\)

Unlike in SVARs, the parameters of the SVMA model have direct economic interpretations as impulse responses. 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 \( j \in \{1, \ldots, n\}, \) choose an \( i_j \in \{1, \ldots, n\} \) and normalize the impact response of variable \( i_j \) to shock \( j: \Theta_{ij,j,0} = 1. \) Then the parameter \( \Theta_{ij,\ell} \) is the expected response at horizon \( \ell \) of variable \( i \) to shock \( j, \) where the size of the shock is of a magnitude that raises variable \( i_j \) by one unit on impact:\(^{11}\)

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

\(^9\)I briefly discuss nonstationarity, nonlinearity, and time-varying parameters in Section 8.

\(^{10}\)See Leeper, Sims & Zha (1996, pp. 6–15) and Sims & Zha (2006, p. 252). They emphasize that the assumption of mutually unpredictable shocks deliberately departs from standard practice in classical linear simultaneous equation models due to the different interpretation of the error terms.

\(^{11}\)Henceforth, moments of the data and shocks are implicitly conditioned on the parameters \((\Theta, \sigma).\)
The impulse response function (IRF) of variable $i$ to shock $j$ is the $(q+1)$-dimensional vector $(\Theta_{ij,0}, \Theta_{ij,1}, ..., \Theta_{ij,q})'$. 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 $\varepsilon_{j,t}$.

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 $\varepsilon_{2,t}$. The figure imposes 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. As the figure shows, the impulse response parameters $\Theta_{ij,\ell}$ can be visualized jointly in a format that is familiar from theoretical macro modeling. Each impulse response (the crosses in the figure) corresponds to a distinct IRF parameter $\Theta_{ij,\ell}$. In contrast, the parameters in the SVAR model are only 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).**

$$\varepsilon_t \overset{i.i.d.}{\sim} N(0, \text{diag}(\sigma^2_1, \ldots, \sigma^2_n)), \ t \in \mathbb{Z}. \quad (5)$$

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 6 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 robust to violations of the Gaussianity assumption. Moreover, the method for sampling from the posterior in Section 3 is readily adapted to non-Gaussian and/or heteroskedastic likelihoods, as discussed in Section 8.

### 2.3 Invertibility

One advantage of the SVMA model is that it allows for noninvertible IRFs. These can arise in applications in which the econometrician does not observe all variables in economic agents’ information sets. Here I review the prevalence of noninvertible IRFs in macroeconomics and the SVAR model’s inability to consistently estimate such IRFs. Because the SVMA model does not restrict IRFs to be invertible *a priori*, it is applicable to a broader set of empirical settings 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 $(y_t, y_{t-1}, \ldots)$ of the observed data, given knowledge of the parameters.\(^{12}\) In this sense, noninvertibility is caused by economically important variables being omitted from the econometrician’s information set.\(^{13}\) 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. See Hansen &

\(^{12}\)Precisely, the IRFs are invertible if $\varepsilon_t$ lies in the closed linear span of $(y_t, y_{t-1}, \ldots)$. Invertible MA representations are also referred to as “fundamental” in the literature.

Sargent (1981, 1991) and Lippi & Reichlin (1994) for extensive mathematical discussions of invertibility in SVMAs and SVARs.

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.\(^\text{14}\) 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 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,\(^\text{15}\) while others rely on the availability of proxy variables for the shocks, thus ameliorating the invertibility issue.\(^\text{16}\) 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). In fact, the IRFs are invertible if and only if the polynomial \(z \mapsto \det(\Theta(z))\) has no roots inside the unit circle.\(^\text{17}\) In general, the structural

---

\(^\text{14}\)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).

\(^\text{15}\)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.

\(^\text{16}\)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, especially 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.

\(^\text{17}\)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).
shocks can be recovered from past, current, and future values of the observed data:\textsuperscript{18}

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

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 \(u_{t+\ell|t-1} = y_{t+\ell} - \text{proj}(y_{t-1}, y_{t-2}, \ldots)\) 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.

For illustration, consider a univariate SVMA model with \(n = q = 1\):

\[
y_t = \varepsilon_t + \Theta_1 \varepsilon_{t-1}, \quad \Theta_1 \in \mathbb{R}, \quad E(\varepsilon_t^2) = \sigma^2.
\] (6)

If \(|\Theta_1| \leq 1\), the IRF \(\Theta = (1, \Theta_1)\) is invertible: The shock has the SVAR representation \(\varepsilon_t = \sum_{\ell=0}^{\infty} (-\Theta_1)^{\ell} y_{t-\ell}\), so it can be recovered using current and past values of the observed data. On the other hand, 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. Clearly, the latter possibility is fully consistent with the SVMA model (6) but inconsistent with any SVAR model of the form (2):\textsuperscript{19}

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, cf. Appendix A.3.1 and Hansen & Sargent (1981). Moreover, standard state-space methods can always be used to estimate the structural shocks, as demonstrated in Section 5. 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

As in SVAR analysis, the IRFs in the SVMA model are only partially identified. The lack of identification arises because the model treats all shocks symmetrically and because

\textsuperscript{18}See for example 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 \(\det(\Theta(1)) = \det(\sum_{\ell=0}^{\infty} \Theta_{\ell}) = 0\).

\textsuperscript{19}If \(|\Theta_1| > 1\), an SVAR (with \(m = \infty\)) applied to the time series (6) estimates the incorrect invertible IRF \((1, 1/\Theta_1)\) and (Wold) “shock” \(u_{t|t-1} = \varepsilon_t + (1 - \Theta_1^2) \sum_{\ell=1}^{\infty} (-\Theta_1)^{-\ell} \varepsilon_{t-\ell}\). This is because the SVMA parameters \((\Theta_1, \sigma)\) and \((1/\Theta_1, \sigma \Theta_1)\) are observationally equivalent, cf. Section 2.4.
noninvertible IRFs are not ruled out \textit{a priori}.

Because of the linearity of the SVMA model and the assumption of Gaussian shocks, any two IRFs that give rise to the same autocovariance function (ACF) are observationally equivalent. Under Assumption 1, the matrix ACF of the time series \( \{y_t\} \) is given by

\[
\Gamma(k) = E(y_{t+k}y_t') = \begin{cases} 
\sum_{\ell=0}^{q-k} \Theta_{\ell+k} \text{diag}(\sigma)^2 \Theta'_\ell & \text{if } 0 \leq k \leq q, \\
0 & \text{if } k > q. 
\end{cases} \tag{7}
\]

Under Assumptions 1 and 2, the observed vector time series \( y_t \) is a mean-zero strictly stationary Gaussian process, so the distribution of the data is completely characterized by the ACF \( \Gamma(\cdot) \). The identified set \( S \) for the IRF parameters \( \Theta = (\Theta_0, \Theta_1, \ldots, \Theta_q) \) and shock standard deviation parameters \( \sigma = (\sigma_1, \ldots, \sigma_n)' \) is thus a function of the ACF:

\[
S(\Gamma) = \{(\tilde{\Theta}_0, \ldots, \tilde{\Theta}_q) \in \Xi_{\Theta}, \tilde{\sigma} \in \Xi_{\sigma}: \sum_{\ell=0}^{q-k} \tilde{\Theta}_{\ell+k} \text{diag}(\tilde{\sigma})^2 \tilde{\Theta}'_\ell = \Gamma(k), 0 \leq k \leq q \},
\]

where \( \Xi_{\Theta} = \{(\tilde{\Theta}_0, \ldots, \tilde{\Theta}_q) \in \mathbb{R}^{n \times n(q+1)}: \tilde{\Theta}_{ij,0} = 1, 1 \leq j \leq n \} \) is the parameter space for \( \Theta \), and \( \Xi_{\sigma} = \{(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_n)' \in \mathbb{R}^n: \tilde{\sigma}_j > 0, 1 \leq j \leq n \} \) is the parameter space for \( \sigma \).\(^{20}\) 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. Building on Hansen & Sargent (1981) and Lippi & Reichlin (1994), Appendix A.2 provides a constructive characterization of \( S(\Gamma) \). I summarize the main insights here.\(^{21}\) 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 \( S(\Gamma) \) 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 offer no way of distinguishing the first shock from the second shock, say, and consequently no way of separately identifying the IRFs to the first and second shocks.\(^{20}\)

\(^{20}\)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 6.3.

\(^{21}\)The identification problem is not easily cast in the framework of interval identification, as \( S(\Gamma) \) is of strictly lower dimension than the parameter space \( \Xi_{\Theta} \times \Xi_{\sigma} \). Still, expression (7) for \( \text{diag}(\Gamma(0)) \) implies that the identified set for scaled impulse responses \( \Psi_{ij,\ell} = \Theta_{ij,\ell} \sigma_j \) is bounded.
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.

Mathematically, 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 \( \tilde{\Theta} \text{diag}(\tilde{\sigma}) Q = \Theta \text{diag}(\sigma) \).

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 correct 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 (drawn with a thick line) is the one that would be estimated by a SVAR (with lag length \( m = \infty \)). However, there exist three other IRFs that have very different economic implications but are equally consistent with the observed ACF.\(^{22}\) If \( n > 1 \), the identification problem is even more severe, as described in Appendix A.2.

As the data alone does not suffice to distinguish between IRFs that have very different economic implications but are equally consistent with the observed ACF.\(^{22}\) If \( n > 1 \), the identification problem is even more severe, as described in Appendix A.2.

\(^{22}\)Similarly, in the special case \( n = q = 1 \), the parameters \((\Theta_1, \sigma)\) imply 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 \( |\Theta_1| \leq 1 \), 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 \( \varepsilon_{t-1} \).
economic implications, it is necessary to leverage additional prior information. In SVAR analysis the prior information is often referred to as the identification scheme, cf. Section 7. The next subsection describes the flexible and transparent approach to prior specification I adopt for the SVMA model.

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 through a transparent and flexible Bayesian prior distribution. Researchers often have access to more prior information about IRFs than what SVAR methods exploit. I explain how such information helps distinguish between observationally equivalent IRFs. Then I propose a prior elicitation procedure that imposes all types of prior information about IRFs in a unified way. I highlight a Gaussian prior family that is convenient to visualize, but as Gaussianity is not essential for my approach, I discuss other choices of priors as well.

To impose prior information, the researcher must have some knowledge about the identity and effects of the unobserved shocks. As in the SVAR approach, 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, must be imposed. Prior information can be imposed dogmatically (with 100% certainty, as is common in SVAR analysis) or non-dogmatically (with less than 100% certainty).

Types and sources of prior information. Because the SVMA model is parametrized in terms of IRFs, it is possible to exploit many types of prior information. Researchers often have fairly weak – but not agnostic – 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 percent. Researchers typically have more informative priors about the signs of certain impulse responses, e.g., the impact response of the output gap to a monetary policy shock that raises the federal funds rate. Researchers may also have quite informative 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

---

23 Kline & Tamer (2015) develop methods for conducting Bayesian inference about the identified set in general models. Unfortunately, as argued above, hypotheses that only concern the identified set $S(\Gamma)$ are rarely interesting in the context of estimating structural impulse responses $\Theta_{ij,\ell}$ because such hypotheses must treat all types of shocks symmetrically.

24 The order of the shocks is immaterial.
peters out). Finally, researchers often have strong beliefs about the smoothness of IRFs, due to adjustment costs, time to build, 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 microeconometric 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 5. 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 $\Theta$. In this approach an informative joint prior distribution is placed 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 place a flexible multivariate prior distribution on the IRFs and shock standard deviations. The generality of this approach necessitates the use of simulation methods for conducting posterior inference about the structural parameters. The simulation method I propose in Section 3 works with any prior distribution for which the log density and its gradient can be computed, giving the researcher great flexibility.

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 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 SVMA analysis thus depends crucially on the prior information imposed, just as SVAR analysis depends on the identification scheme. The frequentist asymptotics in Section 6 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, but it does

\footnote{Alternatively, one could specify a prior on the ACF $\Gamma$ and a conditional prior for $(\Theta, \sigma)$ given $\Gamma$. This approach has the conceptual advantage that the data asymptotically dominates the prior for $\Gamma$ but does not provide information about $(\Theta, \sigma)$ given $\Gamma$ (cf. Section 6). However, in applications, prior information typically directly concerns the IRFs $\Theta$, and it is unclear how to select a meaningful prior for $\Theta$ given $\Gamma$.}

\footnote{From a subjectivist Bayesian perspective, as long as the prior is a proper probability distribution, the validity of posterior inference is unaffected by the under-identification of the parameters (Poirier, 1998).}
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.

underscore the need for a transparent and flexible prior elicitation procedure.

Gaussian prior. While many priors are possible, I first discuss an especially convenient multivariate Gaussian prior distribution. The assumption of Gaussianity means that 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 joint prior distribution on the impulse responses is given by

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

\[ \text{Corr}(\Theta_{ij,\ell+k}, \Theta_{ij,\ell}) = \rho_{ij}^k, \quad 0 \leq \ell \leq \ell + k \leq q, \] (8)

for each \((i,j)\). This correlation structure means that the prior smoothness of IRF \((i,j)\) is governed by \(\rho_{ij}\), as illustrated below.\(^{27}\) For simplicity, the IRFs \((\Theta_{ij,0}, \Theta_{ij,1}, \ldots, \Theta_{ij,q})\) are \textit{a priori} independent across \((i,j)\) pairs. The normalized impulse responses have \(\mu_{i,j,0} = 1\) and \(\tau_{i,j,0} = 0\) for each \(j\). The shock standard deviations \(\sigma_1, \ldots, \sigma_n\) are \textit{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 \(j\).\(^{28}\) In practice, the prior variances \((\tau_j^\sigma)^2\) for the log shock standard deviations

\(^{27}\)The prior has the equivalent autoregressive representation \((\Theta_{ij,\ell+1} - \mu_{ij,\ell+1})/\tau_{ij,\ell+1} = \rho_{ij}(\Theta_{ij,\ell} - \mu_{ij,\ell})/\tau_{ij,\ell} + (1 - \rho_{ij}^2)\zeta_{ij,\ell+1}\), where \(\zeta_{ij,\ell}\) is i.i.d. \(N(0,1)\). That is, if the true impulse response at horizon \(\ell\) is above its prior mean, then we also find it likely that the true impulse response at horizon \(\ell + 1\) is above its prior mean, and more likely the higher \(\rho_{ij}\) is.

\(^{28}\)Alternatively, the prior on \(\sigma\) could be derived from a prior on the forecast error variance decomposition, cf. definition (9) in Section 5. I leave this possibility to future research.
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 $\tau_{ij,\ell}^2$ 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.\(^{29}\)

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 suitable values for the $\mu_{ij,\ell}$ and $\tau_{ij,\ell}^2$ hyperparameters.\(^{30}\) 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 features such as the sign and magnitude of the impact response of the output gap to a monetary policy shock.

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 7, SVAR analyses exploit zero or sign restrictions on individual impulse responses or linear combinations thereof, while information about the shapes and smoothness of IRFs is neglected. Furthermore, prior restrictions on short- or long-run responses implicitly restrict other features of the IRFs, since

---

\(^{29}\)In principle, the $\rho_{ij}$ values could be chosen by an empirical Bayes method. For each possible choice of $\rho_{ij}$, one could compute the marginal likelihood of the data (Chib, 2001, Sec. 10.2) and select the value of $\rho_{ij}$ that maximizes the marginal likelihood. I leave this possibility to future research.

\(^{30}\)The prior confidence bands in Figures 4 and 5 are pointwise bands that consider each horizon separately. This is the most common way to express uncertainty about impulse responses. Sims & Zha (1999, Sec. 6) recommend quantifying uncertainty about entire impulse response functions, i.e., uniform bands.
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 $(i, j)$. 
the VAR model structure subtly constrains the possible shapes of the IRFs.

Bayesian analysis in the SVMA model is explicit about the prior restrictions on IRFs, and researchers can draw on standard Bayesian tools for conducting sensitivity analysis and model validation. The entire set of prior beliefs about IRFs is easily expressed graphically, unlike in SVAR analysis. The sensitivity of posterior inference with respect to features of the prior can be assessed using tools from the comprehensive Bayesian literature (Lopes & Tobias, 2011; Müller, 2012). Model validation and comparison can be carried out through the flexible framework of prior and posterior predictive checks and computation of Bayes factors.\textsuperscript{31} I give examples of prior sensitivity and model validation checks in the empirical application in Section 5. In contrast, robustness checks in SVAR analyses are typically limited to considering a small set of alternative identifying restrictions.

\textbf{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 (8), or to impose heavy-tailed or asymmetric prior distributions on certain impulse responses. Section 7 gives examples of priors that transparently impose well-known identifying restrictions from the SVAR literature.

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 very cumbersome.\textsuperscript{32} 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. If desired, a reweighting step can undo the Whittle approximation at the end.

I first define the posterior density of the structural parameters. Let $T$ be the sample

\textsuperscript{31}See Chib (2001, Ch. 10), Geweke (2010, Ch. 2), and Gelman et al. (2013, Ch. 6).

\textsuperscript{32}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.
size and $Y_T = (y'_1, y'_2, \ldots, y'_T)'$ 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 $\Psi = (\Psi_0, \Psi_1, \ldots, \Psi_q)$, where $\Psi_\ell = \Theta_\ell \text{diag}(\sigma)$ for $\ell = 0, 1, \ldots, q$. 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 draw from the posterior distribution, I use a variant of MCMC known as Hamiltonian Monte Carlo (HMC). HMC is known to offer superior performance over other generic MCMC methods when the dimension of the parameter vector is high. 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.

HMC outperforms standard Random Walk Metropolis-Hastings algorithms because it exploits information contained in the gradient of the log posterior density to systematically explore the posterior distribution. See Neal (2011) for a very readable overview of HMC. 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 $(\Theta^{(1)}, \sigma^{(1)}), \ldots, (\Theta^{(N)}, \sigma^{(N)})$, each of which is (very nearly) distributed according to the posterior distribution. 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 (Chib, 2001, pp. 3579, 3596). 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.

---

33Gelman et al. (2013, Ch. 11) discuss methods for checking that the chain has converged.
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 simulation study in the next section a typical 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, cf. Appendix A.3.1, but a faster alternative is to use the Whittle (1953) approximation to the likelihood of a stationary Gaussian process. Appendix A.3.2 shows that both the Whittle log likelihood and the Whittle score for the SVMA model can be calculated efficiently using the Fast Fourier Transform. 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 Appendix A.4.1. As initial value for the HMC iterations I use a rough approximation to the posterior mode obtained using the constructive characterization of the identified set in Appendix A.2. The HMC algorithm I use adapts to the posterior standard deviations of individual parameters in a warm-up phase; this speeds up computations in some applications. Matlab code for implementing the full inference procedure is available on my website, cf. Footnote 1.

**Reweighting.** Appendix A.4.2 describes an optional reweighting step that translates the Whittle HMC draws into draws from the exact posterior \( p_{\Theta,\sigma|Y}(\Theta, \sigma \mid Y_T) \). If the HMC algorithm is run with the Whittle likelihood and score replacing the exact likelihood and score, the algorithm yields draws from the “Whittle posterior” density \( p_{W|Y}(\Theta, \sigma \mid Y_T) \propto p_{W|Y}(Y_T \mid \Psi(\Theta, \sigma))p_{\Theta,\sigma}(\Theta, \sigma) \), where \( p_{W|Y}(Y_T \mid \Psi(\Theta, \sigma)) \) is the Whittle likelihood. Reweighting can be used if the researcher seeks finite-sample optimal inference under the Gaussian SVMA model.

---

34Hansen & 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. Moreover, the asymptotic theory in Qu & Tkachenko (2012b) assumes identification, unlike Section 6.

35The exact score can be approximated using finite differences, but this is highly time-consuming. The Koopman & Shephard (1992) analytical score formula is not applicable here due to the singular measurement density in the state-space representation of the SVMA model, cf. Appendix A.3.1.
The reweighting step is fast and does not require computation of score vectors.

The asymptotic analysis in Section 6.3 shows that the reweighting step has negligible effect in large samples, as the exact and Whittle posteriors converge to the same limit under weak nonparametric conditions. However, in applications where the MA lag length $q$ is large relative to the sample size, the asymptotic distribution may not be a good approximation to the finite-sample posterior, and reweighting may have a non-negligible effect.

4 Simulation study

To illustrate the workings of the SVMA approach, I conduct a small simulation study with two observed variables and two shocks. I show that prior information about the smoothness of the IRFs can substantially sharpen posterior inference. It is thus desirable to use an approach, like the SVMA approach, for which prior information about smoothness is directly controlled. I also illustrate the consequences of misspecifying the prior.

The illustration is based on the bivariate example from Section 2 with $n = 2$ and $q = 10$, cf. Figure 1. The number of parameters is $n^2(q+1) = 2^2(10+1) = 44$, smaller than the dimensionality of realistic empirical applications but sufficient to elucidate the flexibility, transparency, and effectiveness of the SVAR approach.

Parameters and prior. I consider a single parametrization, with a prior that is correctly centered but diffuse. The sample size is $T = 200$. The true IRF parameters $\Theta$ are the noninvertible ones plotted in Figure 1. The true shock standard deviations are $\sigma_1 = 1$ (monetary policy shock) and $\sigma_2 = 0.5$ (demand shock). I first show results for the prior specification in Figure 4 with $\rho_{ij} = 0.9$ for all $(i,j)$. The prior is centered at the true values but it expresses significant prior uncertainty about the magnitudes of the individual impulse responses. The prior on $\sigma = (\sigma_1, \sigma_2)$ is median-unbiased for the true values but it is very diffuse, with prior standard deviation of $\log \sigma_j$ equal to $\tau_j^{\sigma} = 2$ for $j = 1, 2$.

Simulation settings. I simulate a single sample of artificial data from the Gaussian SVMA model and then run the HMC algorithm using the Whittle likelihood (I do not reweight the draws as in Appendix A.4.2). I take 10,000 MCMC steps, storing every 10th step and discarding the first 3,000 steps as burn-in.\footnote{The results are virtually identical in simulations with 100,000 MCMC steps.} The full computation takes less than 3
hours in Matlab 8.6 on a personal laptop with 2.3 GHz Intel CPU. Appendix A.5.1 provides graphical diagnostics on the convergence and mixing of the MCMC chain.

Baseline results. Figure 6 shows that the posterior for the IRFs accurately estimates the true values and that the data serves to substantially reduce the prior uncertainty. The posterior means are generally close to the truth, although the means for two of the IRFs are slightly too low in this simulation. The 5–95 percentile posterior credible intervals are mostly much narrower than the prior 90% confidence bands, so this prior specification successfully allows the researcher to learn from the data about the magnitudes of the impulse responses. Figure 7 shows the posterior draws for the shock standard deviations and compares them with the prior distribution. The posterior draws are centered around the true values despite the very diffuse prior on $\sigma$. Overall, the inference method for this choice of prior works well, despite the noninvertibility of the true IRFs.

Role of prior smoothness. To illustrate the importance of prior information about the smoothness of the IRFs, I run the HMC algorithm with the same specification as above,
Figure 7: Summary of posterior shock standard deviation ($\sigma$) draws for the bivariate SVMA model with prior smoothness $\rho_{ij} = 0.9$. The plots show the true value (thick vertical line), prior density (curve), and histogram of posterior draws, for each $\sigma_j$, $j = 1, 2$.

Figure 8: Summary of posterior IRF ($\Theta$) draws for the bivariate SVMA model with prior smoothness $\rho_{ij} = 0.3$. See caption for Figure 6.
except that I set \( \rho_{ij} = 0.3 \) for all \((i,j)\) in the prior, as in Figure 5. Figure 8 summarizes the posterior distribution of the IRFs corresponding to this alternative prior. Compared to Figure 4, the posterior credible intervals are much wider and the posterior means are less accurate estimates of the true IRFs.

The higher the degree of prior smoothness, the more do nearby impulse responses “learn from each other”. Due to the prior correlation structure (8), any feature of the data that is informative about the impulse response \( \Theta_{ij,\ell} \) is also informative about \( \Theta_{ij,\ell+k} \); more so for smaller values of \(|k|\), and more so for larger values of the smoothness hyperparameter \( \rho_{ij} \). Hence, a higher degree of prior smoothness reduces the effective number of free parameters in the model. If the true IRFs are not smooth but the prior imposes a lot of smoothness, posterior inference can be very inaccurate. It is therefore important to use a framework, like the SVMA approach, where prior smoothness is naturally parametrized and directly controlled. SVAR IRFs also impose smoothness *a priori*, but the degree of smoothness is implicitly controlled by the VAR lag length and restrictions on the VAR coefficients.

**Misspecified priors.** Appendix A.5.2 reports results for modifications of the baseline simulation above, maintaining the prior distribution but substantially modifying the true IRFs. I consider two such experiments: one in which the shocks have less persistent effects than the prior indicates, and one in which the true IRF of the output gap to a monetary policy shock is uniformly zero. In both cases, the inaccurate prior is overruled by the data, delivering reasonably accurate posterior inference. This happens because the implied prior distribution of the ACF is inconsistent with the true ACF. Since the data is informative about the latter, the posterior distribution puts more weight than the prior on parameters that are consistent with the true ACF, as shown formally in Section 6.3.\(^{37}\)

5 Application: News shocks and business cycles

I now use the SVMA method to infer the role of technological news shocks in the post-war 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 partially informed by a conventional sticky-price DSGE model, without imposing the model restrictions dogmatically.

\(^{37}\text{By the same token, if the true parameter values were chosen to be observationally equivalent to the prior medians in Figure 4 (i.e., they imply the same ACF), then the posterior would look the same as in Figures 6 and 7 up to simulation noise, even though the true parameters could be very different from the prior medians. Hence, not all misspecified priors can be corrected by the data, cf. Section 6.3.}\)
The analysis finds overwhelming evidence of noninvertible IRFs in my specification, yet most of the IRFs are estimated precisely. Furthermore, news shocks are relatively unimportant drivers of productivity and output growth, but more important for the real interest rate. Graphical diagnostics show that the posterior inference is insensitive to moderate changes in the prior; they also point to possible fruitful extensions of the model.

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. My news shock analysis is the first to fully allow for noninvertible IRFs while refraining from dogmatically imposing a particular DSGE model structure (see the discussion at the end of this section).

**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 $T = 213$ (a quarter is lost when transforming to growth rates). I exclude data from 2008 to the present as my analysis ignores financial shocks.

TFP growth equals 100 times the log growth rate of TFP and is taken from the data appendix to Fernald (2014).\footnote{The TFP measure is based on a growth accounting method that adjusts for differing marginal products of capital across sectors as well as changes over time in labor quality and labor’s share of income. Fernald (2014) also estimates utilization-adjusted TFP, but the adjustment is model-based and reliant on estimates from annual regressions on a separate dataset, so I prefer the simpler series. Data downloaded July 14, 2015.} The remaining data is from the St. Louis Federal Reserve’s FRED database.\footnote{FRED series codes: A939RX0Q048SBEA (real GDP per capita), FEDFUNDS (effective federal funds rate), and IPDNBS (implicit price deflator, non-farm business sector). Data downloaded August 13, 2015.} Real GDP growth is given by 100 times the log growth rate of seasonally adjusted GDP per capita in chained dollars, as measured by the Bureau of Economic Analysis (NIPA Table 7.1, line 10). My real interest rate series equals the nominal policy interest rate minus the contemporaneous inflation rate.\footnote{If agents form inflation expectations under the presumption that quarterly inflation follows a random walk, then my measure of the real interest rate equals the conventional *ex ante* real interest rate.} The nominal policy rate is the average effective federal funds rate, expressed as a quarterly rate. The inflation rate equals 100 times the log growth rate in the seasonally adjusted implicit price deflator for the non-farm business sector, as reported by the Bureau of Labor Statistics.
Figure 9: Raw data on TFP growth, GDP growth, and the real interest rate (IR), along with estimated time-varying trends (smooth curves). The final data used in the empirical analysis are differences between the raw series and the trends.

Before running the analysis, I detrend the three data series to remove secular level changes that are arguably unrelated to the business cycle. Following Stock & Watson (2012, Sec. I.C), I estimate the trend in each series using a biweight kernel smoother with a bandwidth of 100 quarters; the trends are then subtracted from the raw series. Figure 9 plots the raw data and the estimated time-varying trends.

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 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 10 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, internal habit formation in consumption, and interest rate smoothing in the Taylor rule. 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.41

As prior means for the nine SVMA IRFs I use the corresponding IRFs implied by the log-

---

41 My approach is distinct from IRF matching (Rotemberg & Woodford, 1997). In IRF matching, a SVAR is identified using exclusion restrictions, and then the structural parameters of a DSGE model are chosen so that the DSGE-implied IRFs match the estimated SVAR IRFs. In my procedure, the DSGE model informs the choice of prior on IRFs, but then the data is allowed to speak through a flexible SVMA model. I do not treat the DSGE model as truth, and I impose prior restrictions in a single stage. Ingram & Whiteman (1994) and Del Negro & Schorfheide (2004) apply related ideas to VAR models. Geweke (2010, Ch. 4.4) proposes a general method for letting DSGE models inform priors without imposing model restrictions dogmatically.
linearized DSGE model, with one exception mentioned below. 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 (not necessarily zero) 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 stickiness in the economy 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 quite 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 based on economic intuition and 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 $\mu_{\sigma_j}$ of $\log(\sigma_j)$ is set to $\log(0.5)$, while the prior standard deviation $\tau_{\sigma_j}$ is set to 2. These values should of course depend on the units of the observed series.

As a consistency check, Appendix A.6.1 shows that the Bayesian computation procedure

---

42The DSGE-implied IRFs for the real interest rate use the same definition of this variable as in the construction of the data series, i.e., nominal interest rate minus contemporaneous inflation. The IRFs are computed using Dynare 4.4.3 (Adjemian et al., 2011).

43A theoretically more satisfying way to deal with uncertainty about the anticipation horizon is to use a Gaussian mixture prior, where the categorical component label is the anticipation horizon.

44The prior is agnostic about the relative importance of the three shocks: Due to the diffuse prior on shock standard deviations, unreported simulations show that the prior 5th and 95th percentiles of the FEVD (cf. (9)) are very close to 0 and 1, respectively, for almost all $(i, j, \ell)$ combinations.
with the above prior accurately recovers the DSGE-implied IRFs from simulated data.

**Results.** Given my prior, the data is informative about most of the IRFs. Figure 11 summarizes the posterior distribution of each impulse response. Figure 12 plots the posterior distribution of long-run (i.e., cumulative) impulse responses $\sum_{\ell=0}^q \Theta_{ij,\ell}$ for each variable-shock combination $(i, j)$. Figure 13 shows the posterior distribution of the forecast error variance decomposition (FEVD) of each variable $i$ to each shock $j$ at each horizon $\ell$, defined as

\[
FEVD_{ij,\ell} = \frac{\text{Var}(\sum_{k=0}^q \Theta_{ij,k} \varepsilon_{j,t+k} | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)}{\text{Var}(y_{i,t+\ell} | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)} = \frac{\sum_{k=0}^\ell \Theta_{ij,k}^2 \sigma_j^2}{\sum_{b=1}^n \sum_{k=0}^\ell \Theta_{ib,k}^2 \sigma_b^2}.
\]

$FEVD_{ij,\ell}$ 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 the shocks up to time $t - 1$.

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. Figure 12 shows that the monetary policy shock has negative and substantially non-neutral effects on the level of GDP in the long run, even though the prior distribution for this long-run response is centered around zero.

The IRFs corresponding to the news shock are not estimated as precisely as IRFs for the other shocks, but the data does noticeably update the prior. 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 data, 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 is found to have substantially less persistent effects on GDP growth than predicted by the DSGE model. However, the effect of the news shock on the real interest rate is found to be large and persistent.

The news shock is not an important driver of TFP and GDP growth but is important

---

45 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 $(y_{t-1}, y_{t-2}, \ldots)$ instead of $(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$. This distinction matters when the IRFs are noninvertible. Baumeister & Hamilton (2015a) conduct inference on the FEVD in a Bayesian SVAR, assuming invertibility.
Figure 11: Summary of posterior IRF (θ) draws, news shock application. The plots show prior 90% confidence bands (shaded), posterior means (crosses), and posterior 5–95 percentile intervals (vertical bars).

Figure 12: Histograms of posterior draws of long-run impulse responses $\sum_{\ell=0}^{q} \Theta_{ij,\ell}$ for each $(i,j)$, news shock application. Curves are prior densities. Histograms and curves each integrate to 1.
for explaining real interest rate movements at longer horizons. According to Figure 13, 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 in explaining GDP growth, while the productivity shock is much more important by these measures. However, the monetary 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 14 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 $R^2$ from a population regression of each shock $\varepsilon_{j,t}$ on current and 50 lags of past data $(y_{t}, y_{t-1}, \ldots, y_{t-50})$, assuming i.i.d. Gaussian shocks.\footnote{Given the parameters, I run the Kalman filter in Appendix A.3.1 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 $\text{Var}(\text{diag}(\sigma)^{-1} \varepsilon_{51} \mid y_{51}, \ldots, y_{t})$ whose diagonal elements equal 1 minus the desired population $R^2$ values at the given parameters.} This $R^2$ value should be essentially 1 for all shocks if the IRFs are invertible, by definition. Instead, Figure 14 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.\footnote{Essentially no posterior IRF draws are exactly invertible; the prior probability is 0.06%}. 

---

**Figure 13**: Summary of posterior draws of $FEVD_{i,j,\ell}$ (9), 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 $j$. 

---

\[ $R^2$ value should be essentially 1 for all shocks if the IRFs are invertible, by definition. Instead, Figure 14 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.\footnote{Essentially no posterior IRF draws are exactly invertible; the prior probability is 0.06%}. 

---

46Given the parameters, I run the Kalman filter in Appendix A.3.1 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 $\text{Var}(\text{diag}(\sigma)^{-1} \varepsilon_{51} \mid y_{51}, \ldots, y_{t})$ whose diagonal elements equal 1 minus the desired population $R^2$ values at the given parameters. 

47Essentially no posterior IRF draws are exactly invertible; the prior probability is 0.06%. 

---

34
Figure 14: 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.

Figure 15: Posterior distribution of the invertible IRFs that are closest to the actual IRFs, news shock application. The figure shows posterior means of actual IRFs from Figure 11 (thick lines), posterior means of the closest invertible IRFs (crosses), and posterior 5–95 percentile intervals for these invertible IRFs (vertical bars).
The noninvertibility of the estimated IRFs is economically significant. Figure 15 summarizes the posterior distribution of those invertible IRFs that are closest to the actual (possibly noninvertible) IRFs. Specifically, for each posterior draw \((\Theta, \sigma)\) I compute the parameter vector \((\tilde{\Theta}, \tilde{\sigma})\) that minimizes the Frobenius distance \(\|\Theta \text{diag}(\sigma) - \tilde{\Theta} \text{diag}(\tilde{\sigma})\|\) over parameters for which \(\tilde{\Theta}\) is invertible and \((\tilde{\Theta}, \tilde{\sigma})\) generates the same ACF as \((\Theta, \sigma)\).\(^{48}\) While the invertible IRFs for the productivity and monetary policy shocks are similar to the unrestricted IRFs, the invertible news shock IRFs look nothing like the actual estimated IRFs.\(^{49}\) Thus, no SVAR identification scheme can deliver accurate inference about the effects of technological news shocks in this dataset.

Appendix A.6.2 uses standard state-space methods to estimate the structural shocks, which is straight-forward despite noninvertibility of the IRFs.

**Prior sensitivity and model validation.** In Appendix A.6.3 I show that the posterior inference is insensitive to moderate changes in the prior distribution. I use the Müller (2012) measure of local prior sensitivity, which allows me to graphically summarize the sensitivity of the posterior mean of each impulse response.

I conduct a battery of graphical posterior predictive checks to identify ways to improve the model’s fit. As shown in Section 6, the posterior distribution of the parameters of the Gaussian SVMA model fits the unconditional second moments of the observed data well in large samples. In Appendix A.6.4 I investigate whether the model also matches higher moments and conditional time series properties. While the Gaussianity-based posterior analysis is robust to violations of Gaussianity and other model assumptions, cf. Section 6, the posterior predictive analysis points to ways the model could be improved to increase statistical efficiency. The analysis suggests it would be fruitful to extend the model to include stochastic volatility and nonlinearities. I briefly discuss such extensions in Section 8.

**Comparison with the literature.** My conclusion that technological news shocks are not important for explaining business cycles is consistent with the literature, but my method is the first to allow for noninvertibility without additional assumptions. Forni et al. (2014)

\(^{48}\)According to Appendix A.2, \((\tilde{\Theta}, \tilde{\sigma})\) is obtained as follows. First apply transformation (ii) in Theorem 3 several times to \((\Theta, \sigma)\) in order to flip all roots outside the unit circle. Denote the resulting invertible parameters by \((\tilde{\Theta}, \tilde{\sigma})\). Then \(\tilde{\Theta} \text{diag}(\tilde{\sigma}) = \tilde{\Theta} \text{diag}(\tilde{\sigma})Q\), where \(Q\) is the orthogonal matrix that minimizes \(\|\Theta \text{diag}(\sigma) - \tilde{\Theta} \text{diag}(\tilde{\sigma})Q\|\). This is an “orthogonal Procrustes problem”, whose solution is well known.

\(^{49}\)Figure 15 cannot be interpreted as the posterior distribution corresponding to a prior which truncates the prior from Figure 10 to the invertible region. It is difficult to sample from this truncated posterior, as essentially none of the unrestricted posterior draws are invertible, so an accept-reject scheme is inapplicable.
estimate small effects of technological news shocks in a factor-augmented SVAR. Their empirical strategy may overcome the noninvertibility issue if technological news are well captured by the first few principal components of their large macroeconomic panel data set. They confirm that low-dimensional systems (without factors) are noninvertible. Papers that estimate fully-specified DSGE models with news shocks also tend to find a limited role for technological news, cf. the review by Beaudry & Portier (2014, Sec. 4.2.2). Unlike these papers, I do not dogmatically impose restrictions implied by a particular structural model.

Several SVAR papers on news shocks have used stock market data in an attempt to overcome the invertibility problem, cf. Beaudry & Portier (2014, Sec. 3). Such SVAR specifications may be valid if the stock market is a good proxy for the news shock, i.e., if the market responds immediately and forcefully upon arrival of technological news. On the other hand, if market movements are highly contaminated by other types of shocks, incorporating stock market data may lead to biased SVAR estimates. It would be interesting to incorporate stock market data into my analysis to fully reconcile my results with these SVAR analyses.

6 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 that includes the SVMA model. 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.

6.1 General results for partially identified models

In this subsection I present a general result on the frequentist asymptotic limit of the Bayesian posterior distribution for a large class of partially identified models that includes the SVMA model. Due to the lack of identification, the asymptotic analysis is nonstandard, as the data does not dominate all aspects of the prior distribution 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 $\Gamma$. 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 and equipped with the Frobenius norm $\| \cdot \|$.

As an illustration, consider the SVMA model with data vector $Y_T = (y'_1, \ldots, y'_T)'$. Let $\Gamma = (\Gamma(0), \ldots, \Gamma(q))$ 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., $\theta = (\Theta_{11,0}, \ldots, \Theta_{11,q})'$. 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 $Y_T \perp \perp \theta | \Gamma$.

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

$$P_{\theta|Y}(A | Y_T) = \int_{\Xi_\Gamma} P_{\theta|\Gamma}(A | \Gamma) P_{\Gamma|Y}(d\Gamma | Y_T)$$

for any measurable set $A \subset \Xi_\theta$. Let $\Gamma_0$ denote the true value of $\Gamma$. If the reduced-form parameter $\Gamma_0$ is identified, the posterior $P_{\Gamma|Y}(\cdot | Y_T)$ for $\Gamma$ will typically concentrate around $\Gamma_0$ in large samples, so that the posterior for $\theta$ is well approximated by $P_{\theta|Y}(\cdot | Y_T) \approx P_{\theta|\Gamma}(\cdot | \Gamma_0)$, 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 measurable functions $h(\cdot)$ bounded in absolute value by 1.

**Lemma 1.** Let the posterior measure $P_{\theta|Y}(\cdot | Y_T)$ satisfy the decomposition (10). All stochastic limits below are taken under the true probability measure of the data. Assume:

---

50 $T$ denotes the sample size, but the model does not have to be a time series model.

51 The $L_1$ distance $\| P_1 - P_2 \|_{L_1}$ between two probability measures $P_1$ and $P_2$ equals twice the total variation distance (TVD) between $P_1$ and $P_2$. TVD is an important metric, as convergence in TVD implies convergence of Bayes point estimators under certain side conditions (van der Vaart, 1998, Ch. 10.3).
(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 $\| \cdot \|_{L_1}$. \(^{52}\)

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

Then as $T \rightarrow \infty$,

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

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

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

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. \(^{53}\) 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 $\theta$ lies in a lower-dimensional subspace of $\Xi_{\theta}$. \(^{54}\)

If continuity of $\Pi_{\theta|\Gamma}(\cdot \mid \Gamma)$ does not hold, assumption (ii) on the limiting posterior distribution

---

\(^{52}\)Denote the underlying probability sample space by $\Omega$, and let $B_{\theta}$ be the Borel sigma-algebra on $\Xi_{\theta}$. Formally, assumption (i) requires the existence of a function $\varsigma : B_{\theta} \times \Xi_{\Gamma} \rightarrow \mathbb{R}_+$ such that $\{\varsigma(B, \Gamma(o))\}_{B \in B_{\theta}, o \in \Omega}$ is a version of the conditional probability measure of $\theta$ given $\Gamma$, and such that $\|\varsigma(\cdot, \Gamma_k) - \varsigma(\cdot, \Gamma_0)\|_{L_1} \rightarrow 0$ as $k \rightarrow \infty$ for any sequence $\{\Gamma_k\}_{k \geq 1}$ satisfying $\Gamma_k \rightarrow \Gamma_0$ and $\Gamma_k \in \Xi_{\Gamma}$.

\(^{53}\)As emphasized by Gustafson (2015, pp. 35, 59–61), the Bayesian approach to partial identification explicitly acknowledges the role of prior information even in infinite samples. This stands in contrast with traditional “identified” models, for which the potential bias due to misspecification of the identifying restrictions is often unacknowledged and difficult to characterize.

\(^{54}\)For a discussion of this point, see Remarks 2 and 3, pp. 768–770, in Moon & Schorfheide (2012).
for $\Gamma$ can be strengthened to derive an asymptotic approximation to the posterior for $\theta$, cf. Moon & Schorfheide (2012, Sec. 3).

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 on $\Gamma$. As long as the reduced-form parameter $\Gamma_0$ is identified, posterior consistency holds under very weak regularity conditions, as discussed in Appendix A.7.1 and in the next subsection.\footnote{See also Ghosh & Ramamoorthi (2003, Ch. 1.3) and Choudhuri, Ghosal & Roy (2005, Ch. 3).}

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 $\tilde{\Gamma} \mapsto p_{Y|\Gamma}(Y_T | \tilde{\Gamma})$ denotes the likelihood function for $\Gamma$ used to compute the posterior $P_{\Gamma|Y}(\cdot | Y_T)$, then $p_{Y|\Gamma}(Y_T | \Gamma_0)$ need not be the true density of the data. As long as $P_{\Gamma|Y}(\cdot | Y_T)$ is a probability measure that satisfies assumption (ii), where the convergence in probability occurs under the true probability measure of the data, then the conclusion of the lemma follows. This observation is helpful when I derive the limit of the Whittle posterior for the SVMA model.

### 6.2 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, as in the SVMA model. The result below supposes that the posterior measure for the ACF $\Gamma$ is computed using the Whittle likelihood under the working assumption that the time series is stationary Gaussian and $q$-dependent, i.e., the autocovariances after lag $q$ are zero. This is the case for the SVMA model. I show that the Whittle posterior is consistent for the true ACF (up to lag $q$) even if the true data generating process is in fact not Gaussian or $q$-dependent.\footnote{In the case of i.i.d. data, posterior consistency in misspecified models has been investigated in detail, see Ramamoorthi, Sriram & Martin (2015) and references therein. Shalizi (2009) places high-level assumptions on the prior and likelihood to derive posterior consistency under misspecification with dependent data. Müller (2013) discusses decision theoretic properties of Bayes estimators when the model is mispecified. Tamaki (2008) derives a large-sample Gaussian approximation to the Whittle-based posterior under a correctly specified parametric spectral density and further regularity conditions.}

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

(iv) For any fixed integer \( k \geq 0 \), \[ T^{-1} \sum_{t=k+1}^{T} y_t y'_{t-k} \xrightarrow{p} \Gamma_0(k) \text{ as } T \to \infty. \]

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.\(^{57}\)

To state the posterior consistency result, I first define the posterior measure. Let \[ \mathbb{T}_{n,q} = \left\{ \{\Gamma(k)\}_{0 \leq k \leq q} : \Gamma(\cdot) \in \mathbb{R}^{n \times n}, \ \Gamma(0) = \Gamma(0)', \right\} \]

be the space of ACFs for \(n\)-dimensional full-rank nondeterministic \(q\)-dependent processes. Let \( p^W_{Y|\Gamma}(Y_T | \Gamma) \) denote the Whittle approximation to the likelihood of a stationary Gaussian process with ACF \( \Gamma.\(^{58}\) Let \( \Pi(\cdot) \) be a prior measure on the space \( \mathbb{T}_{n,q} \). The associated Whittle posterior measure for \( \{\Gamma_0(k)\}_{0 \leq k \leq q} \) given the data \( Y_T \) is given by

\[ P^W_{\Gamma|Y}(A | Y_T) = \frac{\int_A \int_{\mathbb{T}_{n,q}} \frac{p^W_{Y|\Gamma}(Y_T | \Gamma)\Pi(\Gamma)(d\Gamma)}{p^W_{Y|\Gamma}(Y_T | \Gamma)\Pi(\Gamma)(d\Gamma)}}{\int_{\mathbb{T}_{n,q}} \frac{p^W_{Y|\Gamma}(Y_T | \Gamma)\Pi(\Gamma)(d\Gamma)}{p^W_{Y|\Gamma}(Y_T | \Gamma)\Pi(\Gamma)(d\Gamma)}}, \] (11)

where \( A \) is a measurable subset of \( \mathbb{T}_{n,q} \).

\(^{57}\)Phillips & Solo (1992) and Davidson (1994, Ch. 19) give sufficient conditions for the law of large numbers for dependent data.

\(^{58}\)The precise functional form is stated in Appendix A.7.2.
Theorem 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 \( U \) of \( \{ \Gamma_0(k) \}_{0 \leq k \leq q} \) in \( T_{n,q} \), we have
\[
P_{\Pi_Y}^{W}(U \mid Y_T) \overset{p}{\rightarrow} 1,
\]
as \( T \to \infty \) under the true probability measure of the data.

The SVMA model (3) and (5) is an example of a model with a stationary Gaussian and \( q \)-dependent likelihood. Hence, when applied to the SVMA model, Theorem 1 states that if the prior measure on the SVMA parameters induces a prior measure on \( \Gamma \) which has the true ACF \( \{ \Gamma_0(k) \}_{0 \leq k \leq q} \) in its support, then the model-implied Whittle posterior for \( \Gamma \) precisely pins down the true ACF in large samples. This result is exploited in the next subsection.

While the measure \( P_{\Pi_Y}^{W}(A \mid Y_T) \) is computed using the Whittle likelihood and therefore exploits the working assumption that the data is Gaussian and \( q \)-dependent, Theorem 1 shows that posterior consistency for the true ACF (up to lag \( q \)) holds even for time series that are not Gaussian or \( q \)-dependent. The only restrictions placed on the true distribution of the data are the stationarity and weak dependence conditions in Assumption 3. Theorem 1 is silent about posterior inference on autocovariances at lags higher than \( q \), although the true higher-order autocovariances are allowed to be nonzero.

Theorem 1 places no restrictions on the prior \( \Pi_{\Gamma}(\cdot) \) on the ACF, except that the true ACF \( \Gamma_0 \) lies in its support. This level of generality is helpful below when I derive the properties of the SVMA posterior, since no closed-form expression is available for the prior on the ACF that is induced by any given prior on the IRFs and shock standard deviations.

The intermediate results I derive in Appendix A.7 to prove Theorem 1 may be useful in other contexts. My proof of Theorem 1 is based on the general Lemma 3 in Appendix A.7.1, which gives sufficient conditions for posterior consistency in any model, time series or otherwise. Another component of the proof is Lemma 4, which builds on Dunsmuir & Hannan (1976) to show posterior consistency for the reduced-form (Wold) IRFs in an invertible MA model with \( q \) lags, where the posterior is computed using the Whittle likelihood, but the data only has to satisfy Assumption 3. No assumptions are placed on the prior, except that the true reduced-form IRFs must be contained in its support.

6.3 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. That is, \( \theta = \{\Theta_{i,\ell}\}_{1 \leq i \leq n, 0 \leq \ell \leq q}, \sigma_1 \) and \( \Gamma = (\Gamma(0), \ldots, \Gamma(k)) \). I now apply Lemma 1 and Theorem 1 to the SVMA model, which will give a simple description of the limiting form of the Whittle posterior \( P_W^{\theta|Y}(\cdot|Y_T) \) 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,

\[
\Gamma(k) = \sigma_1^2 \sum_{\ell=0}^{q-k} \Theta_{1,\ell+k} \Theta'_{1,\ell} + \sum_{j=2}^{n} \sigma_j^2 \sum_{\ell=0}^{q-k} \Theta_{j,\ell+k} \Theta'_{j,\ell}, \quad k = 0, 1, \ldots, q, \tag{12}
\]

where \( \Theta_{j,\ell} = (\Theta_{1j,\ell}, \ldots, \Theta_{nj,\ell})' \). If \( \theta = \{\Theta_{i,\ell}\}_{1 \leq i \leq n, 0 \leq \ell \leq q}, \sigma_1 \) and there are two or more shocks \( (n \geq 2) \), then the above equations for \( k = 0, 1, \ldots, q \) 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 \( U \) is continuously distributed, we can expect the conditional prior distribution of \( \theta \) given \( \Gamma \) to be continuous in \( \Gamma \).

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. (12), and so continuity of the conditional prior \( \Pi_{\theta|\Gamma}(\cdot|\Gamma) \) 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 theorem below states the limiting form of the Whittle posterior under general

---

59 This paragraph is inspired by Remark 3, pp. 769–770, in Moon & Schorfheide (2012).
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 (7), this prior induces a joint prior measure \( \Pi_{\Theta,\sigma,\Gamma}(\cdot) \) on \( \Xi_\Theta \times \Xi_\sigma \times \Xi_\Gamma \), 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 \( P_{W_Y}^W(\cdot \mid Y_T) \) denote the Whittle posterior measure for \( \theta \) computed using the Whittle SVMA likelihood, cf. Section 3, and the prior \( \Pi_{\Theta,\sigma}(\cdot) \).

**Theorem 2.** Let the data \( Y_T = (y'_1, \ldots, y'_T)' \) 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

\[
\| P_{\theta|Y}^W(\cdot \mid Y_T) - \Pi_{\theta|\Gamma}(\cdot \mid \Gamma_0) \|_{L_1} \overset{p}{\to} 0,
\]

where the stochastic limit is taken as \( T \to \infty \) under the true probability measure of the data. If \( \hat{\Gamma} = \{\hat{\Gamma}(k)\}_{0 \leq k \leq q} \) denotes the sample autocovariances up to order \( q \), then the above two convergence statements also hold with \( \Gamma_0 \) replaced by \( \hat{\Gamma} \).

Continuity of the conditional prior \( \Pi_{\theta|\Gamma}(\cdot \mid \Gamma) \) is stated as a high-level assumption in Theorem 2. I conjecture that prior continuity holds for the multivariate Gaussian prior introduced in Section 2.5, for the reasons discussed below equation (12), but I have not yet been able to prove this result formally.

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.

---

60 The proof of Theorem 2 follows easily from Lemma 1 and Theorem 1. \( P_{\theta|Y}^W(\cdot \mid Y_T) \) satisfies the general decomposition (10) for partially identified models, where the Whittle posterior for \( \Gamma \) has the general form (11) for \( q \)-dependent Gaussian time series. Theorem 1 gives posterior consistency for \( \Gamma_0 \), which is assumption (ii) in Lemma 1. Posterior consistency for \( \Gamma_0 \) requires the induced prior measure \( \Pi_{\Gamma}(\cdot) \) to have \( \Gamma_0 \) in its support, which is guaranteed by the assumption of full support for the prior \( \Pi_{\Theta,\sigma}(\cdot) \).

61 I conjecture that my results can be extended to allow for the asymptotic thought experiment \( q = q(T) = O(T^\nu) \), for appropriate \( \nu > 0 \) and under additional nonparametric conditions.
How the data updates the prior. According to Theorem 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 a given set of 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 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. In this sense, the SVMA approach lets the data influence IRF inference more flexibly than SVAR analysis, although the degree to which the data influences the posterior depends on the prior.

Theorem 2 shows to what extent the data can falsify the prior beliefs. 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 Appendix A.5.2. On the other hand, if the prior distribution on IRFs is tightly concentrated around parameters $(\Theta, \sigma)$ that lie in the true identified set $S(\Gamma_0)$, then the posterior also concentrates around $(\Theta, \sigma)$, regardless of how close $(\Theta, \sigma)$ are to the true structural parameters.

Robustness to misspecified likelihood. Theorem 2 states that the posterior measure, which is 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. The only restrictions on the true data generating process are the stationarity and weak dependence conditions in Assumption 3. Of course,

---

62The local projection method of Jordà (2005) shares this feature but assumes that shocks are observed.

63As a tool for prior elicitation, prior predictive checks can be used to gauge whether the induced prior distribution on the ACF is inconsistent with the observed sample ACF.

64Baumeister & Hamilton (2015b) derive an analogous result for Bayesian inference in the SVAR model with a particular family of prior distributions and assuming invertibility.
the IRF parameters only have a structural economic interpretation if the basic SVMA model structure in Assumption 1 holds. In this case, the ACF has the form (7), 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).\footnote{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 $\hat{\Gamma}$.}

The asymptotic robustness to non-Gaussianity of the shocks 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, the sampling distribution is sensitive to departures from Gaussianity, but this sensitivity does not matter for the first-order asymptotic limit of the posterior for $\theta$.

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 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.\footnote{Consider the trivial SVMA model $y_t = \varepsilon_t$, $E(\varepsilon_t^2) = \sigma^2 (n = 1, q = 0)$. It is well known that the Gaussian MLE $\hat{\sigma}^2 = T^{-1} \sum_i y_i^2$ of $\sigma^2 = \Gamma(0)$ enjoys unique robustness properties.} It is an interesting question how to exploit the identifying power of non-Gaussian shocks without unduly compromising computational tractability or
robustness to misspecification.

**Theorem 2** also implies that the error incurred in using the Whittle approximation to the SVMA likelihood is negligible in large samples, in the sense that the data pins down the true ACF in large samples even when the Whittle approximation is used.\(^{67}\) This is true whether or not the data distribution is the one implied by the Gaussian SVMA model, as long as Assumption 3 holds. As discussed in Section 3, the reweighting step in Appendix A.4.2 therefore makes no difference asymptotically.

## 7 Comparison with SVAR methods

To aid readers who are familiar with SVARs, this section shows that standard SVAR identifying restrictions can be transparently imposed through specific prior choices in the SVMA model, if desired. The SVMA approach easily accommodates exclusion and sign restrictions on short- and long-run impulse responses. External instruments can be exploited in the SVMA framework by expanding the vector of observed time series. Both dogmatic and non-dogmatic prior restrictions are feasible. For extensive discussion of SVAR identification schemes, see Ramey (2015), Stock & Watson (2015), and Uhlig (2015).

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 unchanged.

Another popular type of identifying restrictions are exclusion restrictions on long-run (i.e., cumulative) impulse responses: \(\sum_{\ell=0}^{q-1} \Theta_{ij,\ell} = 0\) 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 \((\Theta_{ij,0}, \ldots, \Theta_{ij,q-1})\) are treated as free parameters whose prior must be specified by the researcher. When evaluating the 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

---

\(^{67}\)For completeness, I am currently working on a formal proof that the convergence in **Theorem 2** also holds for the posterior computed using the exact Gaussian SVMA likelihood.
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.\(^68\) 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 \( \psi \),\(^69\) 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 \textit{dogmatic} prior restrictions that impose exclusion restrictions with 100\% prior certainty, but in many cases \textit{non-dogmatic} restrictions are more credible.\(^70\) 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 \((\Theta_{ij,0}, \ldots, \Theta_{ij,q-1})\), this relationship fully specifies the prior mean vector and covariance matrix of the entire IRF \((\Theta_{ij,0}, \ldots, \Theta_{ij,q})\).

These considerations only concern the functional form of the prior density for \( \Theta \); evaluation of the likelihood and score is carried out exactly as in Section 3.

Many recent 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 simply restricting the IRF parameter space \( \Xi_\Theta \) to the subspace where the inequality constraints hold. This may require some care when running the HMC procedure, but the standard reparametrization \( \Theta_{ij,\ell} = \pm \exp\{\log(\pm \Theta_{ij,\ell})\} \) should work (see also Neal, 2011, Sec. 5.1). If the researcher is uncomfortable imposing much more prior information than the sign restrictions, the prior distribution for the impulse responses in question can be chosen to be diffuse (e.g., truncated Gaussian with large variance).\(^71\)

However, researchers often have more prior information about impulse responses than

\(^{68}\)These restrictions should include the normalizations \( \Theta_{i,j,0} = 1 \) for \( j = 1, \ldots, n \).

\(^{69}\)For example, a prior for \( \psi \) can be elicited as follows. Elicit a tentative multivariate Gaussian prior for \( \Theta \) that is approximately consistent with the linear restrictions. Then obtain the prior for \( \psi \) from the relationship \( \psi = (C^\perp C^\perp)^{-1} C^\perp \{\text{vec}(\Theta) - C'(CC')^{-1}d\} \). In general, subtle issues (the Borel paradox) arise when a restricted prior is obtained from an unrestricted one (Drèze & Richard, 1983, Sec. 1.3).

\(^{70}\)The distinction between dogmatic (exact) and non-dogmatic (“stochastic”) identifying restrictions is familiar from the Bayesian literature on simultaneous equation models (Drèze & Richard, 1983).

\(^{71}\)Poirier (1998, Sec. 4) warns against entirely flat priors in partially identified models.
just their signs, and this can be exploited in the SVMA approach.\footnote{Similar points have been made in the context of SVARs by Kilian & Murphy (2012), and even more explicitly by Baumeister & Hamilton (2015c). While these papers focus on prior information about impact impulse responses, the SVMA approach facilitates imposing information about longer-horizon responses.} If an impulse response is viewed as being likely to be positive, then very small positive values ought to be less likely than somewhat larger values, up to a point. Additionally, extremely large values for the impulse responses can often be ruled out \emph{a priori}. The multivariate Gaussian prior distribution in Section 2.5 is capable of expressing a strong but non-dogmatic prior belief that certain impulse responses have certain signs, while at the same time imposing weak information about magnitudes and ruling out extreme values.\footnote{In some applications, a non-symmetric (e.g., log-normal) prior distribution may better express a strong prior belief in the sign of an impulse response, while imposing weak prior restrictions on the magnitude.}

While computationally attractive, the well-known Uhlig (2005) inference procedure for sign-identified SVARs is less transparent than the SVMA approach. Uhlig (2005) uses a conjugate prior on the reduced-form VAR parameters and a uniform conditional prior for the structural parameters given the reduced-form parameters.\footnote{Using the notation of \textit{Footnote 5} and setting $\Sigma = H \diag(\sigma^2) H'$, the prior for the reduced-form parameters $(A_1, \ldots, A_m, \Sigma)$ is a normal-inverse-Wishart distribution, while the prior for the orthogonal matrix $Q$ given the reduced-form parameters is Haar measure restricted to the space where the sign restrictions hold. See Arias et al. (2014) for a full discussion. The Uhlig (2005) prior implies a particular informative prior distribution on IRFs, which could in principle be imposed in the SVMA model. Doing so is not desirable, as the main justification for the Uhlig (2005) prior is its convenience in SVAR analysis.} Because the prior is mainly chosen for computational convenience, it is not very flexible and does not easily allow for non-dogmatic sign restrictions. Furthermore, Baumeister & Hamilton (2015b) show that the Uhlig (2005) procedure imposes unintended and unacknowledged prior information in addition to the acknowledged sign restrictions.\footnote{Giacomini & Kitagawa (2015) develop a robust Bayes SVAR approach that imposes dogmatic exclusion and sign restrictions without imposing any other identifying restrictions. Their goals are diametrically opposed to mine, since one of the motivations of the SVMA approach is to allow for as many types of prior information as possible, including information about magnitudes, shapes, and smoothness.} In contrast, the SVMA prior is flexible and all restrictions that it imposes can be transparently visualized.

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). If such an instrument is available, it 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 \( j = 2, 3, \ldots, n \). Depending on the application, we may also have reason to believe that the non-impact impulse responses \( \Theta_{1j,\ell} \) are (close to) zero for \( \ell \geq 1 \). Such prior beliefs can be imposed like any other exclusion restrictions.

The SVMA IRFs can be restricted to be invertible, if desired. As explained in Section 2.3, SVARs implicitly assume that the IRFs are invertible, although this is more of a bug than a feature. If, for some reason, the researcher wants to impose invertibility \textit{a priori} in SVMA analysis, simply restrict the IRF parameter space \( \Xi_{\Theta} \) to the invertible subspace 
\[
\{ \Theta : \text{det}(\sum_{\ell=0}^{q} \Theta_{\ell} z^{\ell}) \neq 0 \quad \forall \ z \in \mathbb{C} \text{ s.t. } |z| < 1 \}\). Since this represents a nonlinear constraint on the parameter space, it is easiest to carry out posterior inference by first employing the HMC procedure on the unrestricted parameter space and afterwards discarding all posterior draws of \( \Theta \) that are noninvertible. If the procedure ends up discarding a high fraction of posterior draws, the invertibility restriction is called into question.

8 Topics for future research

I conclude by listing some – primarily technical – avenues for future research.

Many SVAR papers seek to identify the IRFs to a single shock (while allowing for other shocks). It would be interesting to investigate whether prior elicitation and Bayesian computations in the SVMA approach can be simplified in the case of single-shock identification.

Whereas the SVMA model in Section 2 does not restrict the IRFs at all before prior information is imposed, an alternative modeling strategy is to impose a flexible parametric structure on the SVMA IRFs. Each IRF could be parametrized as a polynomial, trigonometric, or spline function of the horizon, say. The likelihood and score formulas would be essentially unchanged, and computational advantages due to reduced dimensionality may outweigh the loss of flexibility in specifying the IRF prior. Parametrizing the IRFs in this way would even permit an infinite MA lag length \( (q = \infty) \), while allowing for noninvertibility.

Following the SVAR literature, I assumed that the number of shocks equals the number of observed variables. Yet the SVMA likelihood and Whittle approximation may be evaluated in essentially the same way if the number of shocks exceeds the number of variables. Hence, given a prior on the IRFs, the same Bayesian computations can be applied. The asymptotic analysis also carries through essentially unchanged. However, the characterization of the

---

76 If \( \text{det}(\Theta_0) = 0 \), the IRFs are noninvertible. If \( \text{det}(\Theta_0) \neq 0 \), the roots of the polynomial \( \text{det}(\sum_{\ell=0}^{q} \Theta_{\ell} z^{\ell}) \) equal the roots of \( \text{det}(I_{n} + \sum_{\ell=1}^{q} \Theta_{0,\ell}^{-1} z^{\ell}) \). The latter roots can be obtained as reciprocals of the eigenvalues of the polynomial’s companion matrix (e.g., Hamilton, 1994, Prop. 10.1).

77 This idea is explored by Hansen & Sargent (1981, p. 44) and Barnichon & Matthes (2015).
identified set in Appendix A.2 is substantially affected by allowing for more shocks than variables. It would be interesting to extend the identification analysis to the general case.

Also in line with the SVAR literature, this paper focused on a linear stationary model with constant parameters and Gaussian shocks. However, the key advantages of the SVMA approach — the natural IRF parametrization and the ability to allow for noninvertibility — do not rely on these specific assumptions. As long as the likelihood can be evaluated, Bayesian computation is possible in principle. Future work could explore the computational feasibility and robustness properties of SVMA inference that explicitly incorporates nonstationarity, deterministic components, nonlinearities, time-varying parameters, non-Gaussian shocks, or stochastic volatility. The Whittle likelihood approximation used in Section 3 may work poorly in non-Gaussian models, but other computational tricks may be available.

Prior knowledge that one of the shocks is more volatile than usual on a known set of dates can improve identification of the IRFs, as in Rigobon (2003). The SVMA model can be extended along these lines by allowing the standard deviation of the shock in question to switch deterministically between two different values, while assuming constancy of all other structural parameters. It is straightforward to modify the Kalman filter in Appendix A.3.1 to compute the corresponding likelihood function.

A very diffuse prior on IRFs leads to a multimodal posterior distribution due to under-identification. In extreme cases multimodality may cause problems for the basic HMC procedure. It may be possible to exploit the constructive characterization of the identified set in Appendix A.2 to extend the algorithm so that it occasionally jumps between approximate modes of the posterior. In case of multimodality, posterior uncertainty should be summarized by highest posterior density intervals instead of equal-tailed (e.g., 5–95 percentile) intervals.

Finally, it may be possible to simplify posterior simulation by leveraging the asymptotic theory in Section 6.3. For example, to do exact inference, an approximation to the asymptotic posterior limit could be used as a proposal distribution for importance sampling.

---

78 In the context of cointegrated time series, an analog of the SVMA approach is to do Bayesian inference on the parameters of the Granger representation.
A Technical appendix

This appendix defines the notation used in the paper and provides discussion and results of a more technical nature.

A.1 Notation

$I_n$ is the $n \times n$ identity matrix. $i$ is the imaginary unit that satisfies $i^2 = -1$. If $a$ is a vector, $\text{diag}(a)$ denotes the diagonal matrix with the elements of $a$ along the diagonal in order. If $A$ is a square matrix, $\text{tr}(A)$ is its trace, $\text{det}(A)$ is its determinant, and $\text{diag}(A)$ is the vector consisting of the diagonal elements in order. For an arbitrary matrix $B$, $B'$ denotes the matrix transpose, $\bar{B}$ denotes the elementwise complex conjugate, $B^* = B'$ 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. If $C$ is a positive semidefinite matrix, $\lambda_{\min}(C)$ is its smallest eigenvalues. If $Q$ is an $n \times n$ matrix, it is said to be orthogonal if it is real and $QQ' = I_n$, while it is said to be unitary if $QQ^* = I_n$. The statement $X \perp \perp Y \mid Z$ means that the random variables $X$ and $Y$ are independent conditional on $Z$. If $\mathcal{K}$ is a set, $\mathcal{K}^c$ denotes its closure and $\mathcal{K}^c$ denotes its complement.

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 $\Theta = (\Theta_0, \ldots, \Theta_q)$ 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. In words, the theorem says that if we start with some set of IRFs $\Theta(z)$ contained in the identified set, then we can obtain all other sets of IRFs in the identified set by applying orthogonal rotations to $\Theta(z)$ and/or by “flipping the roots” of $\Theta(z)$. Only a finite sequence of such operations is necessary to jump from one element of the identified set to any other element of the identified set.

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

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

(i) Set $\tilde{\Psi}(z) = \Psi(z)Q$, where $Q$ is an arbitrary orthogonal $n \times n$ matrix.
(ii) Let $\gamma_1, \ldots, \gamma_r$ \((r \leq nq)\) 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 $(1 - \gamma_k z)$. Call the resulting matrix polynomial $\tilde{\Psi}(z)$. If $\gamma_k$ is real, stop.

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

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

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

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

On the other hand, if $(\tilde{\Theta}, \tilde{\sigma}) \in S(\Gamma)$ is an arbitrary point in the identified set satisfying $\det(\tilde{\Theta}(0)) \neq 0$, then $(\tilde{\Theta}, \tilde{\sigma})$ 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 $\tilde{\Psi}(z)$.

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

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). Hence, Theorem 3 states that any set of structural IRFs that are consistent with a given ACF \( \Gamma(\cdot) \) are obtained by applying transformations (i) and (ii) to the Wold IRFs corresponding to \( \Gamma(\cdot) \).

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. The rule used to compute the matrix \( J \) in transformation (ii) is not important for the theorem; in particular, \( J \) can be the Cholesky factor of \( \tilde{\Psi}(0)\tilde{\Psi}(0)^* \).

3. The only purpose of transformation (a) is to enforce the normalizations \( \Theta_{i,j,0} = 1 \).

4. To simplify the math, the theorem restricts attention to IRFs satisfying \( \det(\Theta(0)) = \det(\Theta_0) \neq 0 \). If \( \det(\Theta_0) = 0 \), there exists a linear combination of \( y_{1,t}, \ldots, y_{n,t} \) that is perfectly predictable based on knowledge of shocks \( \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \) occurring before time \( t \). Hence, in most applications, a reasonable prior for \( \Theta \) ought to assign zero probability to the event \( \det(\Theta_0) = 0 \).

5. 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 \( \Psi(z) = \Theta(z) \text{diag}(\sigma) \) 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).

### A.3 Likelihood evaluation

This subsection provides formulas for computing the exact Gaussian likelihood as well as the Whittle likelihood and score for the SVMA model.

#### A.3.1 Exact likelihood via the Kalman filter

Let \( \Psi = \Theta \text{diag}(\sigma) \). The state space representation of the SVMA model is

\[
y_{i,t} = \Psi_i \alpha_t, \quad i = 1, \ldots, n, \ t = 1, \ldots, T,
\]

\[
\alpha_t = \begin{pmatrix} 0 & 0 \\ I_{nq} & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} \tilde{\varepsilon}_t \\ 0 \end{pmatrix}, \quad \tilde{\varepsilon}_t \sim i.i.d. N(0, I_n), \quad t = 2, 3, \ldots, T,
\]

\[
\alpha_1 \sim N(0, I_{n(q+1)}),
\]

54
where $\Psi_i$ is the $n(q+1)$-dimensional $i$-th row vector of $\Psi = (\Psi_0, \Psi_1, \ldots, \Psi_q)$, $\tilde{\varepsilon}_t$ is the $n$-dimensional standardized structural shock vector (each element has variance 1), and $\alpha_t = (\tilde{\varepsilon}'_t, \varepsilon'_{t-1}, \ldots, \varepsilon'_{t-q})'$ is the $n(q+1)$-dimensional state vector.

I use the “univariate treatment of multivariate series” Kalman filter in Durbin & Koopman (2012, Ch. 6.4), since that algorithm avoids inverting large matrices. For my purposes, the algorithm is as follows.

1. Initialize the state forecast mean $a_{1,1} = 0$ and state forecast variance $Z_{1,1} = I_{n(q+1)}$.
   Set $t = 1$.

2. For each $i = 1, \ldots, n$:
   (a) Compute the forecast error $v_{i,t} = y_{i,t} - \Psi_i a_{i,t}$, forecast variance $\lambda_{i,t} = \Psi_i Z_{i,t} \Psi'_i$, and Kalman gain $g_{i,t} = (1/\lambda_{i,t}) Z_{i,t} \Psi'_i$.
   (b) Compute the log likelihood contribution: $L_{i,t} = -\frac{1}{2} \left( \log \lambda_{i,t} + v_{i,t}^2 / \lambda_{i,t} \right)$.
   (c) Update the state forecast mean: $a_{i+1,t} = a_{i,t} + g_{i,t} v_{i,t}$.
   (d) Update the state forecast variance: $Z_{i+1,t} = Z_{i,t} - \lambda_{i,t} g_{i,t} g'_i t$.

3. Let $\tilde{a}_{n+1,t}$ denote the first $nq$ elements of $a_{n+1,t}$, and let $\tilde{Z}_{n+1,t}$ denote the upper left $nq \times nq$ block of $Z_{n+1,t}$. Set

$$a_{1,t+1} = \begin{pmatrix} 0 \\ \tilde{a}_{n+1,t} \end{pmatrix}, \quad Z_{1,t+1} = \begin{pmatrix} I_n & 0 \\ 0 & \tilde{Z}_{n+1,t} \end{pmatrix}.$$ 

4. If $t = T$, stop. Otherwise, increment $t$ by 1 and go to step 2.

The log likelihood $\log p_{Y|\Psi}(Y_T | \Psi)$ is given by $\sum_{t=1}^T \sum_{i=1}^n L_{i,t}$, up to a constant.

A.3.2 Whittle likelihood

Let $Y_T = (y'_1, y'_2, \ldots, y'_T)'$ be the stacked data vector. 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

$$\log p_{Y|\Psi}(Y_T | \Psi) = -\frac{1}{2} nT \log(2\pi) - \frac{1}{2} \log \det(V(\Psi)) - \frac{1}{2} Y'_T V(\Psi)^{-1} Y_T. \quad (13)$$

This is what the Kalman filter in Appendix A.3.1 computes.
For all $k = 0, 1, 2, \ldots, T - 1$, define the Fourier frequencies $\omega_k = 2\pi k / T$, the discrete Fourier transform (DFT) of the data $\tilde{y}_k = (2\pi T)^{-1/2} \sum_{t=1}^{T} e^{-i \omega_k (t-1)} y_t$, the DFT of the MA parameters $\tilde{\Psi}_k(\Psi) = \sum_{\ell=1}^{q+1} e^{-i \omega_k (\ell-1)} \Psi_{\ell-1}$, and the SVMA spectral density matrix $f_k(\Psi) = (2\pi)^{-1} \tilde{\Psi}_k(\Psi) \tilde{\Psi}_k(\Psi)^*$. Let $\|B\|_{\text{max}} = \max_{ij} |B_{ij}|$ denote the maximum norm of any matrix $B = (B_{ij})$. Due to the block Toeplitz structure of $V(\Psi)$,

$$
\|V(\Psi) - 2\pi \Delta F(\Psi) \Delta^*\|_{\text{max}} = O(T^{-1})
$$

as $T \to \infty$. $\Delta$ is an $nT \times nT$ matrix with $(s, t)$ block equal to $T^{-1/2} e^{i \omega_{s-1}(t-1)} I_n$, so that $\Delta \Delta^* = I_{nT}$. $F(\Psi)$ is a block diagonal $nT \times nT$ matrix with $(s, s)$ block equal to $f_s(\Psi)$.

The Whittle (1953) approximation to the log likelihood (13) is obtained by substituting $V(\Psi) \approx 2\pi \Delta F(\Psi) \Delta^*$. This yields the Whittle log likelihood

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

The 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).

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 $\tilde{C}_k(\Psi) = \sum_{\ell=1}^{T} e^{-i \omega_k (\ell-1)} C_{\ell-1}(\Psi)$. Once $\{C_k(\Psi)\}_{0 \leq k \leq T-1}$ have been computed, $\{\tilde{C}_k(\Psi)\}_{0 \leq k \leq T-1}$ can be computed using the Fast Fourier Transform. Finally, let $\tilde{C}_k(\Psi) = \tilde{C}_{T+k}(\Psi)$ for $k = -1, -2, \ldots, 1 - T$.

**Lemma 2.**

$$
\frac{\partial \log p^W_Y(Y_T \mid \Psi)}{\partial \Psi_{\ell}} = -\sum_{\ell=0}^{q} \text{Re}[\tilde{C}_{T-\ell}(\Psi)] \Psi_{\ell}, \quad \ell = 0, 1, \ldots, q.
$$

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

$$
\frac{\partial \log p^W_Y(Y_T \mid \Psi)}{\partial \Theta_{\ell}} = \frac{\partial \log p^W_Y(Y_T \mid \Psi)}{\partial \Psi_{\ell}} \text{diag}(\sigma), \quad \ell = 0, 1, \ldots, q.
$$

---

79 The result (14) is a straight-forward vector generalization of Brockwell & Davis (1991, Prop. 4.5.2).
80 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 $f_{T-k}(\Psi) = f_k(\Psi)$ for $k = 1, 2, \ldots, T$.
81 Again, computation time can be saved by exploiting $C_{T-k}(\Psi) = \overline{C_k(\Psi)}$ for $k = 1, 2, \ldots, T$. 56
\[ \frac{\partial \log p_{Y|\Psi}(Y_T | \Psi)}{\partial \log \sigma_j} = \sum_{i=1}^{n} \sum_{\ell=0}^{q} \frac{\partial \log p_{Y|\Psi}(Y_T | \Psi)}{\partial \Psi_{ij,\ell}} \Psi_{ij,\ell}, \quad j = 1, 2, \ldots, n. \]

### A.4 Bayesian computation: Algorithms

This subsection details my posterior simulation algorithm and the optional reweighting step that translates Whittle draws into draws from the exact likelihood.

#### A.4.1 Implementation of Hamiltonian Monte Carlo algorithm

I here describe my implementation of the posterior simulation algorithm. First I outline my method for obtaining an initial value. Then I discuss the modifications I make to the Hoffman & Gelman (2014) algorithm. The calculations below require evaluation of the log prior density, its gradient, the log likelihood, and the score. Evaluation of the multivariate Gaussian log prior and its gradient is straightforward; this is also the case for many other choices of priors. Evaluation of the Whittle likelihood and its score is described in Appendix A.3.2. Matlab code can be found on my website, cf. Footnote 1.

**Initial value.** The HMC algorithm produces draws from a Markov Chain whose long-run distribution is the Whittle posterior of the SVMA parameters, regardless of the initial value used for the chain. However, using an initial value near the mode of the posterior distribution can significantly speed up the convergence to the long-run distribution. I approximate the posterior mode using the following computationally cheap procedure:

1. Compute the empirical ACF of the data.

2. Run \( q \) steps of the Innovations Algorithm to obtain an invertible SVMA representation that approximately fits the empirical ACF (Brockwell & Davis, 1991, Prop. 11.4.2).

   Denote these invertible parameters by \((\hat{\Theta}, \hat{\sigma})\).

3. Let \( C \) denote the (finite) set of complex roots of the SVMA polynomial corresponding to \((\hat{\Theta}, \hat{\sigma})\), cf. Theorem 3.

4. For each root \( \gamma_j \) in \( C \) (each complex conjugate pair of roots is treated as one root):

   [82] In principle, the Innovations Algorithm could be run for more than \( q \) steps, but this tends to lead to numerical instability in my trials. The output of the first \( q \) steps is sufficiently accurate in my experience.
(a) Let \((\tilde{\Theta}^{(j)}, \tilde{\sigma}^{(j)})\) denote the result of flipping root \(\gamma_j\), i.e., of applying transformation (ii) in Theorem 3 to \((\hat{\Theta}, \hat{\sigma})\) with this root.

(b) Determine the orthogonal matrix \(Q^{(j)}\) such that \(\tilde{\Theta}^{(j)} \text{diag}(\tilde{\sigma}^{(j)}) Q^{(j)}\) is closest to the prior mean \(E(\Theta \text{diag}(\sigma))\) in Frobenius norm, cf. Footnote 48.

(c) Obtain parameters \((\tilde{\Theta}^{(j)}, \tilde{\sigma}^{(j)})\) such that \(\tilde{\Theta}^{(j)} \text{diag}(\tilde{\sigma}^{(j)}) Q^{(j)} = \tilde{\Theta}^{(j)} \text{diag}(\tilde{\sigma}^{(j)})\), i.e., apply transformation (a) in Theorem 3. Calculate the corresponding value of the prior density \(\pi(\tilde{\Theta}^{(j)}, \tilde{\sigma}^{(j)})\).

5. Let \(\tilde{j} = \arg\max_j \pi(\tilde{\Theta}^{(j)}, \tilde{\sigma}^{(j)})\).

6. If \(\pi(\tilde{\Theta}^{(\tilde{j})}, \tilde{\sigma}^{(\tilde{j})}) \leq \pi(\hat{\Theta}, \hat{\sigma})\), go to Step 7. Otherwise, set \((\hat{\Theta}, \hat{\sigma}) = (\tilde{\Theta}^{(\tilde{j})}, \tilde{\sigma}^{(\tilde{j})})\), remove \(\gamma_j\) (and its complex conjugate) from \(\mathcal{C}\), and go back to Step 4.

7. Let the initial value for the HMC algorithm be the parameter vector of the form \(((1 - x)\tilde{\Theta} + x E(\Theta), (1 - x)\tilde{\sigma} + x E(\sigma))\) that maximizes the posterior density, where \(x\) ranges over the grid \(\{0, 0.01, \ldots, 0.99, 1\}\), and \((E(\Theta), E(\sigma))\) is the prior mean of \((\Theta, \sigma)\).

Step 2 computes a set of invertible parameters that yields a high value of the likelihood. Steps 3–6 find a set of possibly noninvertible parameters that yields a high value of the prior density while being observationally equivalent with the parameters from Step 2 (I use a “greedy” search algorithm since it is computationally prohibitive to consider all combinations of root flips). Because Steps 2–6 lexicographically prioritize maximizing the likelihood over maximizing the prior, Step 7 allows the parameters to shrink toward the prior means.

HMC implementation. I use the HMC variant NUTS from Hoffman & Gelman (2014), which automatically tunes the step size and trajectory length of HMC. See their paper for details on the NUTS algorithm. I downloaded the code from Hoffman’s website. I make two modifications to the basic NUTS algorithm, neither of which are essential, although they do tend to improve the mixing speed of the Markov chain in my trials. These modifications are also used in the NUTS-based statistics software Stan (Stan Development Team, 2015).

First, I allow for step size jittering, i.e., continually drawing a new HMC step size from a uniform distribution over some interval (Neal, 2011, Sec. 5.4.2.2). The jittering is started after the stepsizes have been tuned as described in Hoffman & Gelman (2014, Sec. 3.2). For the applications in this paper, the step size is chosen uniformly at random from the interval \([0.5\hat{\epsilon}, 1.5\hat{\epsilon}]\), where \(\hat{\epsilon}\) is the tuned step size.
Second, I allow for a diagonal “mass matrix”, where the entries along the diagonal are estimates of the posterior standard deviations of the SVMA parameters (Neal, 2011, Sec. 5.4.2.4). I first run the NUTS algorithm for a number of steps with an identity mass matrix. Then I calculate the sample standard deviations of the parameter draws over a window of subsequent steps, after which I update the mass matrix accordingly. I update the mass matrix twice more using windows of increasing length. Finally, I freeze the mass matrix for the remainder of the NUTS algorithm. In this paper, the mass matrix is estimated over steps 300–400, steps 401–600, and steps 601–1000, and it is fixed after step 1000.

A.4.2 Reweighting

An optional reweighting step may be used to translate draws obtained from the Whittle-based HMC algorithm into draws from the exact Gaussian posterior density \( p_{\Theta, \sigma | Y}(\Theta, \sigma | Y_T) \). The Whittle HMC algorithm yields draws \((\Theta(1), \sigma(1)), \ldots, (\Theta(N), \sigma(N))\) (after discarding a burn-in sample) from the Whittle posterior density \( p^W_{\Theta, \sigma | Y}(\Theta, \sigma | Y_T) \). If desired, apply the following reweighting procedure to the Whittle draws:

1. For each Whittle draw \( k = 1, 2, \ldots, N \), compute the relative likelihood weight

\[
w_k = \frac{p_{\Theta, \sigma | Y}(\Theta(k), \sigma(k) | Y_T)}{p^W_{\Theta, \sigma | Y}(\Theta(k), \sigma(k) | Y_T)} = \frac{p_{Y | \Psi}(Y_T | \Psi(\Theta(k), \sigma(k)))}{p^W_{Y | \Psi}(Y_T | \Psi(\Theta(k), \sigma(k)))}.
\]

2. Compute normalized weights \( \tilde{w}_k = w_k / \sum_{b=1}^N w_b, k = 1, \ldots, N \).

3. Draw \( N \) samples \((\tilde{\Theta}(1), \tilde{\sigma}(1)), \ldots, (\tilde{\Theta}(N), \tilde{\sigma}(N))\) from the multinomial distribution with mass points \((\Theta(1), \sigma(1)), \ldots, (\Theta(N), \sigma(N))\) and corresponding probabilities \( \tilde{w}_1, \ldots, \tilde{w}_N \).

Then \((\tilde{\Theta}(1), \tilde{\sigma}(1)), \ldots, (\tilde{\Theta}(N), \tilde{\sigma}(N))\) constitute \( N \) draws from the exact posterior distribution. This reweighting procedure is a Sampling-Importance-Resampling procedure (Rubin, 1988) that uses the Whittle posterior as a proposal distribution. The reweighting step is fast, as it only needs to compute the exact likelihood—not the score—for \( N \) different parameter values, where \( N \) is typically orders of magnitude smaller than the required number of likelihood/score evaluations during the HMC algorithm.

---

83 The sample standard deviations are partially shrunk toward 1 before updating the mass matrix.
Figure 16: MCMC chains for each IRF parameter ($\Theta$) in the $\rho_{ij} = 0.9$ simulations in Section 4. Each jagged line represents a different impulse response parameter (two of them are normalized at 1). The vertical dashed line marks the burn-in time, before which all draws are discarded. The horizontal axes are in units of MCMC steps, not stored draws (every 10th step is stored).

A.5 Simulation study: Additional results

Here I provide diagnostics and additional results relating to the simulations in Section 4.

A.5.1 Diagnostics for simulations

I report diagnostics for the baseline $\rho_{ij} = 0.9$ bivariate simulation, but diagnostics for other specifications in this paper are similar. The average HMC acceptance rate is slightly higher than 0.60, which is the rate targeted by the NUTS algorithm when tuning the HMC step size. The score of the posterior was evaluated about 382,000 times. Figures 16 and 17 show the MCMC chains for the IRF and log shock standard deviation draws. Figures 18 and 19 show the autocorrelation functions of the draws.

A.5.2 Simulations with misspecified priors

I here provide simulation results for two bivariate experiments with substantially misspecified priors. I maintain the same prior on IRFs and shock standard deviations as the $\rho_{ij} = 0.9$
Figure 17: MCMC chains for each log shock standard deviation parameter ($\log \sigma$) in the $\rho_{ij} = 0.9$ simulations in Section 4. See caption for Figure 16.

Figure 18: Autocorrelation functions for HMC draws of each IRF parameter ($\Theta$) in the $\rho_{ij} = 0.9$ simulations in Section 4. Each jagged line represents a different impulse response parameter. Only draws after burn-in were used to computed these figures. The autocorrelation lag is shown on the horizontal axes in units of MCMC steps.
prior in Section 4, cf. Figures 4 and 7. Here, however, I modify the true values of the IRFs so they no longer coincide with the prior means.

**Misspecified persistence.** I first consider an experiment in which the prior overstates the persistence of the shock effects, i.e., the true IRFs die out quicker than indicated by the prior means $\mu_{ij,\ell}$ in Figure 4. The true IRFs are set to $\Theta_{ij,\ell} = c_{ij}e^{-0.25\ell}\mu_{ij,\ell}$ for all $(i,j,\ell)$, where $c_{ij} > 0$ is chosen so that $\max_{\ell} |\Theta_{ij,\ell}| = \max_{\ell} |\mu_{ij,\ell}|$ for each IRF. The true shock standard deviations, the prior ($\rho_{ij} = 0.9$), the sample size, and the HMC settings are exactly as in Section 4. Figure 20 compares these true IRFs to the prior distribution. The figure also summarizes the posterior distribution for the IRFs. The posterior is not perfectly centered but is much closer to the truth than the prior is. Figure 21 shows why this is the case: The prior distribution on $(\Theta, \sigma)$ implies a distribution for auto- and cross-correlations of observed variables that is at odds with the true ACF. Since the data is informative about the ACF, the posterior distribution for IRFs puts higher weight than the prior on IRFs that are consistent with the true auto- and cross-correlations.

**Misspecified cross-correlations.** The second experiment considers a prior that mis-specifies the cross-correlations between the observed variables. I set the true IRFs equal to the prior means in Figure 4, except that the true IRF of the output gap to a monetary policy shock equals zero, i.e., $\Theta_{21,\ell} = 0$ for $0 \leq \ell \leq q$. The true shock standard deviations, the
Figure 20: Summary of posterior IRF ($\Theta$) draws for the bivariate SVMA model with a prior that is too persistent relative to the true parameter values. The plots show true values (thick lines), prior 90% confidence bands (shaded), posterior means (crosses), and posterior 5–95 percentile intervals (vertical bars). The prior means (not shown) are the midpoints of the prior confidence bands, as in Figure 4.

Figure 21: Posterior auto- and cross-correlation draws for the bivariate SVMA model with a prior that misspecifices the persistence of the IRFs. The displays plot draws of $\text{Corr}(y_{i,t}, y_{j,t-k} \mid \Theta, \sigma)$, where $i$ indexes rows, $j$ indexes columns, and $k$ runs along the horizontal axes. The top right display, say, concerns cross-correlations between the FFR and lags of the output gap. The plots show true values (thick lines), prior means (dashed lines) and 5–95 percentile confidence bands (shaded), and posterior means (crosses) and 5–95 percentile intervals (vertical bars).
Figure 22: Summary of posterior IRF (Θ) draws for the bivariate SVMA model with a prior that misspecifies the cross-correlations between variables. See caption for Figure 20.

Figure 23: Posterior autocorrelation draws for the bivariate SVMA model with a prior that misspecifies the cross-correlations between variables. See caption for Figure 21.
prior ($\rho_{ij} = 0.9$), the sample size, and the HMC settings are as above. Figure 22 shows that posterior inference is accurate despite the misspecified prior. Again, Figure 23 demonstrates how the data corrects the prior distribution on auto- and cross-correlations, thus pulling the posterior on IRFs toward the true values (although here the true ACF is not estimated as accurately as in Figure 21).

A.6 Application: Additional results

This subsection presents additional results related to the empirical application in Section 5. First, I show that the SVMA procedure accurately estimates IRFs on simulated data. Second, I demonstrate how the Kalman smoother can be used to draw inference about the shocks. Third, I examine the sensitivity of posterior inference with respect to the choice of prior. Fourth, I assess the model’s fit and suggest ways to improve it.

A.6.1 Consistency check with simulated data

I show that the SVMA approach, with the same prior and HMC settings as in Section 5, can recover the true IRFs when applied to data generated by the log-linearized Sims (2012) DSGE model. I simulate data for the three observed variables from an SVMA model with i.i.d. Gaussian shocks. The true IRFs are those implied by the log-linearized Sims (2012) model (baseline calibration) out to horizon $q = 16$, yielding a noninvertible representation. The true shock standard deviations are set to $\sigma = (0.5, 0.5, 0.5)'$. Note that the prior for the IRF of TFP growth to the news shock is not centered at the true IRF, as explained in Section 5. The sample size is the same as for the actual data ($T = 213$).

Figures 24 and 25 summarize the posterior draws produced by the HMC algorithm when applied to the simulated data set. The posterior means accurately locate the true parameter values. The equal-tailed 90% posterior credible intervals are tightly concentrated around the truth in most cases. In particular, inference about the shock standard deviation parameters is precise despite the very diffuse prior.

A.6.2 Inference about shocks

Figure 26 shows the time series of posterior means for the structural shocks given the real dataset. For each posterior draw of the structural parameters $(\Theta, \sigma)$, I compute $E(\varepsilon_t | \Theta, \sigma, Y_T)$ using the smoothing recursions corresponding to the Gaussian state-space representation in Appendix A.3.1 (Durbin & Koopman, 2012, p. 157), and then I average over
Figure 24: Summary of posterior IRF ($\Theta$) draws, simulated news shock data. See caption for Figure 6.

Figure 25: Summary of posterior shock standard deviation ($\sigma$) draws, simulated news shock data. See caption for Figure 7.
draws. If the structural shocks are in fact non-Gaussian, the smoother still delivers mean-square-error-optimal linear estimates of the shocks. If desired, draws from the full joint posterior distribution of the shocks can be obtained from a simulation smoother (Durbin & Koopman, 2012, Ch. 4.9). It is also straightforward to draw from the predictive distribution of future values of the data using standard methods for state-space models.

A.6.3 Prior sensitivity

To gauge the robustness of posterior inference with respect to the choice of prior, I compute the sensitivity measure “PS” of Müller (2012). This measure captures the first-order approximate effect on the posterior means of changing the prior mean hyperparameters. Let \( \theta \) denote the vector containing all impulse responses and log shock standard deviations of the SVMA model, and let \( e_k \) denote the \( k \)-th unit vector. Because my prior for \( \theta \) is a member of an exponential family, the Müller (2012) PS measure for parameter \( \theta_k \) equals

\[
PS_k = \max_{\nu} \frac{\partial E(\theta_k | Y_T)}{\partial \nu} = \sqrt{\epsilon_k \text{Var}(\theta | Y_T) \text{Var}(\theta)^{-1} \text{Var}(\theta | Y_T) e_k}. \tag{16}
\]
Figure 27: $PS_k$ measure of the sensitivity of the posterior IRF means with respect to changes in the prior means of all parameters, cf. (16), in the news shock application. The symmetric vertical bars have length $2PS_k$ and are centered around the corresponding posterior means (crosses).

This is the largest (local) change that can be induced in the posterior mean of $\theta_k$ from changing the prior means of the components of $\theta$ by the multivariate equivalent of 1 prior standard deviation. $PS_k$ depends only on the prior and posterior variance matrices $\text{Var}(\theta)$ and $\text{Var}(\theta | Y_T)$, which are easily obtained from the HMC output.

Figure 27 plots the posterior means of the impulse responses along with $\pm PS_k$ intervals (where the index $k$ corresponds to the $(i,j,\ell)$ combination for each impulse response). The wider the band around an impulse response, the more sensitive is the posterior mean of that impulse response to (local) changes in the prior. In economic terms, most of the posterior means are seen to be insensitive to changes in the prior means of magnitudes smaller than 1 prior standard deviation. The most prior-sensitive posterior inferences, economically speaking, concern the IRF of GDP growth to a news shock, but large changes in the prior means are necessary to alter the qualitative features of the posterior mean IRF.

---

84 In particular, $PS_k \geq \max_b |\partial E(\theta_k | Y_T)/\partial E(\theta_b)| \sqrt{\text{Var}(\theta_b)}$. Whereas $PS_k$ is a local measure, the effects of large changes in the prior can be evaluated using reweighting (Lopes & Tobias, 2011, Sec. 2.4).
A.6.4 Posterior predictive analysis

I conduct a posterior predictive analysis to identify ways to improve the fit of the Gaussian SVMA model (Geweke, 2010, Ch. 2.4.2). For each posterior parameter draw produced by HMC, I simulate an artificial dataset of sample size $T = 213$ from a Gaussian SVMA model with the given parameters. On each artificial dataset I compute four checking functions. First and second, the skewness and excess kurtosis of each series. Third, the long-run autocorrelation of each series, defined as the Newey-West long-run variance estimator (20 lags) divided by the sample variance. Fourth, I run a reduced-form VAR regression of the three-dimensional data vector $y_t$ on its 8 first lags and a constant; then I compute the first autocorrelation of the squared VAR residuals for each of the three series. The third measure captures persistence, while the fourth measure captures volatility clustering in forecast errors.

Figure 28 shows the distribution of checking function values across simulated datasets, as well as the corresponding checking function values for the actual data. The Gaussian SVMA model does not capture the skewness and kurtosis of GDP growth; essentially, the model does not generate recessions that are sufficiently severe relative to the size of booms. The model...
somewhat undershoots the persistence and kurtosis of the real interest rate. The fourth column suggests that forecast errors for TFP and GDP growth exhibit volatility clustering in the data, which is not captured by the Gaussian SVMA model.

The results point to three fruitful model extensions. First, introducing stochastic volatility in the SVMA model would allow for better fit along the dimensions of kurtosis and forecast error volatility clustering. Second, nonlinearities or skewed shocks could capture the negative skewness of GDP growth. Finally, increasing the MA lag length $q$ would allow the model to better capture the persistence of the real interest rate, although this is not a major concern, as I am primarily interested in shorter-run impulse responses.

A.7 Asymptotic theory: Mathematical details

I here give additional details concerning the frequentist asymptotics of Bayes procedures. First I provide high-level sufficient conditions for posterior consistency. Then I state the functional form for the Whittle ACF likelihood mentioned in Section 6.2. Finally, I prove posterior consistency for the parameters in the Wold decomposition of a $q$-dependent time series, which is useful for proving posterior consistency for the ACF (Theorem 1). I allow for misspecification of the likelihood functions used to compute the various posterior measures.

All stochastic limits below are taken as $T \to \infty$, and all stochastic limits and expectations are understood to be taken under the true probability measure of the data. The abbreviation “w.p.a. 1” means “with probability approaching 1 as $T \to \infty$”.

A.7.1 General conditions for posterior consistency

Following Ghosh & Ramamoorthi (2003, Thm. 1.3.4), I give general sufficient conditions for assumption (ii) of Lemma 1. Let $\Pi_\Gamma(\cdot)$ denote the marginal prior measure for parameter $\Gamma$, with parameter space $\Xi_\Gamma$. Let $p_{Y|\Gamma}(Y_T | \Gamma)$ denote the (possibly misspecified) likelihood function. The posterior measure is given by

$$P_{\Gamma|Y}(A \mid Y_T) = \frac{\int_A p_{Y|\Gamma}(Y_T | \Gamma) \Pi_\Gamma(d\Gamma)}{\int_{\Xi_\Gamma} p_{Y|\Gamma}(Y_T | \Gamma) \Pi_\Gamma(d\Gamma)}$$

for measurable sets $A \subset \Xi_\Gamma$.\footnote{I assume throughout the paper that integrals in the definitions of posterior measures are well-defined.}

Lemma 3. Define the normalized log likelihood ratio $\hat{\phi}(\Gamma) = T^{-1} \log \frac{p_{Y|\Gamma}(Y_T | \Gamma)}{p_{Y|\Gamma}(Y_T | \Gamma_0)}$ for all $\Gamma \in$
Assume there exist a function \( \phi: \Xi_\Gamma \to \mathbb{R} \), a neighborhood \( K \) of \( \Gamma_0 \) in \( \Xi_\Gamma \), and a scalar \( \zeta < 0 \) such that the following conditions hold.

(i) \( \sup_{\Gamma \in K} |\hat{\phi}(\Gamma) - \phi(\Gamma)| \xrightarrow{p} 0 \).

(ii) \( \phi(\Gamma) \) is continuous at \( \Gamma = \Gamma_0 \).

(iii) \( \phi(\Gamma) < 0 \) for all \( \Gamma \neq \Gamma_0 \).

(iv) \( \sup_{\Gamma \in K^c} \hat{\phi}(\Gamma) < \zeta \) w.p.a. 1.

(v) \( \Gamma_0 \) is in the support of \( \Pi_{\Gamma}(\cdot) \).

Then for any neighborhood \( U \) of \( \Gamma_0 \) in \( \Xi_\Gamma \), \( P_{\Gamma|Y(U \mid Y_T)} \xrightarrow{p} 1 \).

**Remarks:**

1. The uniform convergence assumption (i) on the log likelihood ratio can often be obtained from pointwise convergence using stochastic equicontinuity (Andrews, 1992).

2. If the likelihood \( p_{Y|\Gamma}(Y_T \mid \Gamma) \) is correctly specified (i.e., \( p_{Y|\Gamma}(Y_T \mid \Gamma_0) \) is the true density of the data) and assumption (i) holds, \( \phi(\Gamma) \) equals the negative Kullback-Leibler divergence and assumptions (ii)–(iii) will typically be satisfied automatically if \( \Gamma \) is identified (Ghosh & Ramamoorthi, 2003, Ch. 1.2–1.3). Even if the likelihood is misspecified, such as with the use of a Whittle likelihood in time series models, it may still be possible to prove the uniform convergence in assumption (i) to some \( \phi(\cdot) \) function that uniquely identifies the true parameter \( \Gamma_0 \) through assumptions (ii)–(iii), as in Theorem 1. This phenomenon is analogous to the well-known consistency property of quasi maximum likelihood estimators under certain types of misspecification.

**A.7.2 Whittle likelihood for a \( q \)-dependent process**

I now state the functional form for the Whittle ACF likelihoods for \( q \)-dependent processes mentioned in Section 6.2. Define the spectral density for a \( q \)-dependent process parametrized in terms of its ACF:

\[
f(\omega; \Gamma) = \frac{1}{2\pi} \left( \Gamma(0) + \sum_{k=1}^{q} \left\{ e^{-i\omega k} \Gamma(k) + e^{i\omega k} \Gamma(k)' \right\} \right), \quad \Gamma \in \mathbb{T}_{n,q}.
\]
Let $\hat{\Gamma}(k) = T^{-1} \sum_{t=1}^{T-k} y_{t+k} y_t^\prime$, $k = 0, 1, \ldots, T - 1$, be the $k$-th sample autocovariance, and set $\hat{\Gamma}(k) = \hat{\Gamma}(-k)^\prime$ for $k = -1, -2, \ldots, 1 - T$. Define the periodogram

$$\hat{I}(\omega) = \frac{1}{2\pi T} \left( \sum_{t=1}^{T} e^{-it\omega} y_t \right) \left( \sum_{t=1}^{T} e^{it\omega} y_t^\prime \right) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} e^{-ik\omega} \hat{\Gamma}(k), \quad \omega \in [-\pi, \pi].$$

The Whittle ACF log likelihood is given by

$$\log p_{Y|\Gamma}^W(Y_T | \Gamma) = -nT \log(2\pi) - \frac{T}{4\pi} \int_{-\pi}^{\pi} \log \det(f(\omega; \Gamma)) \, d\omega - \frac{T}{4\pi} \int_{-\pi}^{\pi} \text{tr}\{f(\omega; \Gamma)^{-1} \hat{I}(\omega)\} \, d\omega.$$

As in Appendix A.3.2, it is common to use a discretized Whittle log likelihood that replaces integrals of the form $(2\pi)^{-1} \int_{-\pi}^{\pi} g(\omega) \, d\omega$ (for some $2\pi$-periodic function $g(\cdot)$) with corresponding sums $T^{-1} \sum_{k=0}^{T-1} g(\omega_k)$, where $\omega_k = 2\pi k/T$ for $0 \leq k \leq T - 1$. The discretization makes it possible to compute the periodogram from the DFT $\tilde{y}_k$ of the data (see Appendix A.3.2), since $\hat{I}(\omega_k) = \tilde{y}_k \tilde{y}_k^\prime$ for $0 \leq k \leq T - 1$. The proof of Theorem 1 shows that posterior consistency also holds when the discretized Whittle likelihood is used.

**A.7.3 Posterior consistency for Wold parameters**

The proof of Theorem 1 relies on a posterior consistency result for the Wold IRFs and prediction covariance matrix in a MA model with $q$ lags. I state this result below. The posterior for the reduced-form parameters is computed using the Whittle likelihood and thus under the working assumption of a MA model with i.i.d. Gaussian innovations. However, consistency only requires Assumption 3, so the true data distribution need not be Gaussian or $q$-dependent. The result in this subsection concerns (invertible) reduced-form IRFs, not (possibly noninvertible) structural IRFs. While the consistency result may be of general interest, in this paper I use it only as a stepping stone for proving Theorem 1.

**Definitions.** 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 given by $\{\Gamma_0(k)\}_{0 \leq k \leq q} \in \mathbb{T}_{n,q}$, i.e., the true ACF out to lag $q$. That is, $\beta_0 = (\beta_{0,1}, \ldots, \beta_{0,q}) \in \mathbb{B}_{n,q}$ and $\Sigma_0 \in \mathbb{S}_n$ are the unique parameters such that $\Gamma_0(k) = \sum_{\ell=0}^{q-k} \beta_{0,\ell+k} \Sigma_0 \beta_{0,\ell}^\prime$ for $0 \leq k \leq q$, where $\beta_{0,0} = I_n$. Here $\mathbb{S}_n$ denotes the space of symmetric positive definite $n \times n$ matrices, while $\mathbb{B}_{n,q}$ is the space of coefficients for which the MA lag
polynomial $\beta_0(L)$ has all its roots outside the unit circle, i.e.,

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

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

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

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

Using the notation introduced in Appendix A.7.2 for the periodogram $\hat{I}(\omega)$, the Whittle MA log likelihood is given by

$$\log p_{Y \mid \beta, \Sigma}^W(Y_T \mid \beta, \Sigma) = -nT \log(2\pi) - \frac{T}{4\pi} \int_{-\pi}^\pi \log \det(\tilde{f}(\omega; \beta, \Sigma)) \, d\omega$$

$$- \frac{T}{4\pi} \int_{-\pi}^\pi \text{tr}\{\tilde{f}(\omega; \beta, \Sigma)^{-1} \hat{I}(\omega)\} \, d\omega.$$

As shown in the proof, the result below goes through if the integrals in the Whittle likelihood are replaced with discretized sums (cf. Appendix A.7.2).

RESULT. 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 $\mathbb{B}_{n,q} \times S_n$. Define the Whittle posterior measure

$$P_{\beta, \Sigma \mid Y}^W(A \mid Y_T) = \frac{\int_A p_{Y \mid \beta, \Sigma}^W(Y_T \mid \beta, \Sigma) \Pi_{\beta, \Sigma}(d\beta, d\Sigma)}{\int_{\mathbb{B}_{n,q} \times S_n} p_{Y \mid \beta, \Sigma}^W(Y_T \mid \beta, \Sigma) \Pi_{\beta, \Sigma}(d\beta, d\Sigma)}$$

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

**Lemma 4.** Let Assumption 3 hold. Assume that the pseudo-true parameters $(\beta_0, \Sigma_0) \in \mathbb{B}_{n,q} \times S_n$ are in the support of the prior $\Pi_{\beta, \Sigma}(\cdot)$. Then the Whittle posterior for $(\beta_0, \Sigma_0)$ is consistent, i.e., for any neighborhood $\bar{U}$ of $(\beta_0, \Sigma_0)$ in $\mathbb{B}_{n,q} \times S_n$,

$$P_{\beta, \Sigma \mid Y}^W(\bar{U} \mid Y_T) \overset{p}{\rightarrow} 1.$$
B Proofs

This appendix contains proofs of the results stated in Section 6 and Appendices A.2 and A.7. The proofs require knowledge of the notation and concepts introduced in these sections. I first give the proofs regarding the identified set for the SVMA model and the gradient of the SVMA Whittle log likelihood. Then I give the proofs regarding the frequentist asymptotics of Bayes procedures.

B.1 Proof of Theorem 3

As in Lippi & Reichlin (1994, p. 311), define the rational matrix function

\[ R(\gamma, z) = \begin{pmatrix} \frac{z-\gamma}{1-\gamma z} & 0 \\ 0 & I_{n-1} \end{pmatrix}, \quad \gamma, z \in \mathbb{C}. \]

Transformation (ii) corresponds to the transformation \( \tilde{\Psi}(z) = \Psi(z)QR(\gamma_k, z)^{-1} \) if \( \gamma_k \) is real. If \( \gamma_k \) is not real, the transformation corresponds to \( \tilde{\Psi}(z) = \tilde{\Psi}(z)\tilde{Q} \), where \( \tilde{\Psi}(z) = \Psi(z)QR(\gamma_k, z)^{-1}\tilde{Q}R(\gamma_k, z)^{-1} \) and \( \tilde{Q} = \tilde{\Psi}(0)^{-1}J \) is a unitary matrix. I proceed in three steps.

**Step 1.** Consider the first claim of the theorem. Let \( f(\omega; \Gamma) = (2\pi)^{-1} \sum_{k=-q}^{q} \Gamma(k)e^{-ik\omega} \), \( \omega \in [-\pi, \pi] \), denote the spectral density matrix function associated with the ACF \( \Gamma(\cdot) \). Since \( \Psi(z) = \Theta(z) \text{diag}(\sigma) \) with \( (\Theta, \sigma) \in S(\Gamma) \), we must have \( \Psi(e^{-i\omega})\Psi(e^{-i\omega})^* = 2\pi f(\omega; \Gamma) \) for all \( \omega \) by the usual formula for the spectral density of a vector MA process (Brockwell & Davis, 1991, Example 11.8.1). Because \( R(\gamma, e^{-i\omega})R(\gamma, e^{-i\omega})^* = I_n \) for any \( (\gamma, \omega) \), it is easy to verify that \( \tilde{\Psi}(z) \) – constructed by applying transformation (i) or transformation (ii) to \( \Psi(z) \) – also satisfies \( \tilde{\Psi}(e^{-i\omega})\tilde{\Psi}(e^{-i\omega})^* = 2\pi f(\omega; \Gamma) \). Hence, \( \tilde{\Psi}(z) = \sum_{\ell=0}^{q-k} \tilde{\Psi}_{\ell+k}z^\ell \) is a matrix MA polynomial satisfying \( \sum_{\ell=0}^{q-k} \tilde{\Psi}_{\ell+k}\tilde{\Psi}_{\ell}^* = \Gamma(k) \) for all \( k = 0, 1, \ldots, q \). In Step 2 below I show that \( \tilde{\Psi}(z) \) is a matrix polynomial with real coefficients. By construction of \( \tilde{\Theta}(z) = \sum_{\ell=0}^{q} \tilde{\Theta}_{\ell}z^\ell \) and \( \tilde{\sigma} \), we then have \( \sum_{\ell=0}^{q-k} \tilde{\Theta}_{\ell+k} \text{diag}(\tilde{\sigma})^2\tilde{\Theta}_{\ell}^* = \Gamma(k) \) for all \( k = 0, 1, \ldots, q \), so \( (\tilde{\Theta}, \tilde{\sigma}) \in S(\Gamma) \), as claimed.

**Step 2.** I now show that transformation (ii) yields a real matrix polynomial \( \tilde{\Psi}(z) \). This fact was asserted by Lippi & Reichlin (1994, pp. 317–318). I am grateful to Professor Marco Lippi for providing me with the proof arguments for Step 2; all errors are my own.

\( \tilde{\Psi}(z) \) is clearly real if the flipped root \( \gamma_k \) is real (since \( \eta \) and \( Q \) can be chosen to be real in this case), so consider the case where we flip a pair of complex conjugate roots \( \gamma_k \) and \( \overline{\gamma_k} \).
Recall that in this case, \( \hat{\Psi}(z) = \tilde{\Psi}(z)\tilde{Q} \), where \( \tilde{\Psi}(z) = \Psi(z)QR(\gamma_k, z)^{-1}\tilde{Q}R(\gamma_k, z)^{-1} \) and \( \tilde{Q} \) is unitary. It follows from the same arguments as in Step 1 that the complex-valued matrix polynomial \( \tilde{\Psi}(z) = \sum_{\ell=0}^{q} \tilde{\Psi}_\ell z^\ell \) satisfies \( \sum_{\ell=0}^{q-k} \tilde{\Psi}_{\ell+k} \tilde{\Psi}_\ell^* = \Gamma(k) \) for all \( k = 0, 1, \ldots, q \)

Let \( \tilde{\Psi}(z) = \sum_{\ell=0}^{q} \Psi_\ell z^\ell \) denote the matrix polynomial obtained by conjugating the coefficients of the polynomial \( \tilde{\Psi}(z) \). By construction, the roots of \( \det(\tilde{\Psi}(z)) \) are real or appear as complex conjugate pairs, so \( \det(\tilde{\Psi}(z)) \) has the same roots as \( \det(\hat{\Psi}(z)) \). Furthermore, for \( k = 0, 1, \ldots, q \),

\[
\sum_{\ell=0}^{q-k} \tilde{\Psi}_{\ell+k} \tilde{\Psi}_\ell^* = \Gamma(k) = \sum_{\ell=0}^{q-k} \Psi_{\ell+k} \Psi_\ell^*.
\]

By Theorem 3(b) of Lippi & Reichlin (1994), there exists a unitary \( n \times n \) matrix \( \tilde{\tilde{Q}} \) such that \( \tilde{\Psi}(z) = \tilde{\Psi}(z)\tilde{\tilde{Q}} \) for \( z \in \mathbb{R} \). The matrix polynomial \( \tilde{\Psi}(z)\tilde{\Psi}(0)^{-1} \) then has real coefficients.\(^{86}\)

For all \( z \in \mathbb{R} \),

\[
\tilde{\Psi}(z)\tilde{\Psi}(0)^{-1} = \left( \tilde{\Psi}(z)\tilde{Q} \right) \left( \tilde{\Psi}(0)\tilde{Q} \right)^{-1} = \tilde{\Psi}(z)\tilde{\Psi}(0)^{-1} = \overline{\tilde{\Psi}(z)\tilde{\Psi}(0)^{-1}}.
\]

Consequently, with the real matrix \( J \) defined as in the theorem, \( \hat{\Psi}(z) = \tilde{\Psi}(z)\tilde{\Psi}(0)^{-1}J \) is a matrix polynomial with real coefficients. Finally, since \( \tilde{Q} \) is unitary, the matrix

\[
\tilde{\Psi}(0)\tilde{\Psi}(0)^* = \left( \tilde{\Psi}(0)\tilde{\tilde{Q}} \right) \left( \tilde{\Psi}(0)\tilde{\tilde{Q}} \right)^* = \overline{\tilde{\Psi}(0)\tilde{\Psi}(0)^*}
\]

is real, symmetric, and positive definite, so \( J \) is well-defined.

**Step 3.** Finally, I prove the second claim of the theorem. Suppose we have a fixed element \((\hat{\Theta}, \hat{\sigma})\) of the identified set that we want to end up with after transforming the initial element \((\Theta, \sigma)\) appropriately. Define \( \tilde{\Psi}(z) = \hat{\Theta}(z)\text{diag}(\hat{\sigma}) \). Since \((\Theta, \sigma), (\hat{\Theta}, \hat{\sigma}) \in \mathcal{S}(\Gamma)\), the two sets of SVMA parameters correspond to the same spectral density, i.e., \( \Psi(e^{-\omega})\Psi(e^{-\omega})^* = \hat{\Psi}(e^{-\omega})\hat{\Psi}(e^{-\omega})^* \) for all \( \omega \in [-\pi, \pi] \). As in the proof of Theorem 2 in Lippi & Reichlin (1994), we can apply transformation (ii) finitely many times (say, \( b \) times) to \( \Psi(z) \), flipping all the roots that are inside the unit circle, thus ending up with a polynomial

\[
B(z) = \Psi(z)Q_1R(\gamma_{k_1}, z)^{-1} \cdots Q_bR(\gamma_{k_b}, z)^{-1}Q_{b+1}
\]

for which all roots of \( \det(B(z)) \) lie on or outside the unit circle. Likewise, denote the (finitely many) roots of \( \det(\tilde{\Psi}(z)) \) by \( \gamma_k, k = 1, 2, \ldots, \), and apply to \( \tilde{\Psi}(z) \) a finite sequence

\(^{86}\)\( \tilde{\Psi}(0) \) is nonsingular because \( \det(\Psi(0)) \neq 0 \).
of transformation (ii) to arrive at a polynomial
\[
\tilde{B}(z) = \tilde{\Psi}(z)Q_1 R(\gamma_{k_1}, z)^{-1} \cdots Q_b R(\gamma_{k_b}, z)^{-1} \tilde{Q}_{b+1}
\]
for which all roots of \( \det(\tilde{B}(z)) \) lie on or outside the unit circle. Since \( \det(B(z)) \) and \( \det(\tilde{B}(z)) \) have all roots on or outside the unit circle, and we have \( B(e^{-i\omega})B(e^{-i\omega})^* = \tilde{B}(e^{-i\omega})\tilde{B}(e^{-i\omega})^* = 2\pi f(\omega; \Gamma) \) for all \( \omega \), there must exist an orthogonal matrix \( Q \) such that \( \tilde{B}(z) = B(z)Q \) (Lippi & Reichlin, 1994, p. 313; Hannan, 1970, p. 69). Thus,
\[
\tilde{\Psi}(z) = \Psi(z)Q_1 R(\gamma_{k_1}, z)^{-1} \cdots Q_b R(\gamma_{k_b}, z)^{-1}Q_{b+1}Q^*_{b+1}R(\gamma_{k_{b+1}}, z)\tilde{Q}_{b+1}^* \cdots R(\gamma_{k_1}, z)\tilde{Q}_1^*;
\]
and
\[
\det(\tilde{\Psi}(z)) = \det(\Psi(z)) (z - \gamma_{k_1}) \cdots (z - \gamma_{k_b})(1 - \gamma_{k_1}^2z) \cdots (1 - \gamma_{k_b}^2z),
\]
so any root \( \gamma_k \) of \( \det(\tilde{\Psi}(z)) \) must either equal \( \gamma_k \) or it must equal \( 1/\gamma_k \), where \( \gamma_k \) is some root of \( \det(\Psi(z)) \). It follows that we can apply a finite sequence of transformation (ii) (i.e., an appropriate sequence of root flips) to \( \Psi(z) \) to obtain a real matrix polynomial \( \tilde{\Psi}(z) \) satisfying \( \det(\tilde{\Psi}(z)) = \det(\tilde{\Psi}(z)) \) for all \( z \in \mathbb{C} \). Theorem 3(b) in Lippi & Reichlin (1994) then implies that \( \tilde{\Psi}(z) \) can be obtained from \( \tilde{\Psi}(z) \) through transformation (i) (i.e., an orthogonal rotation, which clearly must be real). Finally, obtain \( (\tilde{\Theta}, \tilde{\sigma}) \) from \( \tilde{\Psi}(z) \) by transformation (a). \( \square \)

### B.2 Proof of Lemma 2

Suppressing the arguments (\( \Psi \)), let
\[
L_k = \log \det(f_k) + \tilde{\gamma}_k^{-1} \tilde{y}_k.
\]
Then
\[
\frac{\partial L_k}{\partial (f_k')} = f_k^{-1} - f_k^{-1} \tilde{y}_k \tilde{y}_k^* f_k^{-1} = C_k.
\]
Writing \( f_k' = \overline{\Psi}_k \tilde{\Psi}_k' \), we have
\[
\frac{\partial \text{vec}(f_k')}{\partial \text{vec}(\Psi')} = (\tilde{\Psi}_k \otimes I_n) \tilde{e}_{\omega k \ell} + (I_n \otimes \overline{\Psi}_k) K_n e^{-i\omega k \ell},
\]
where \( K_n \) is the \( n^2 \times n^2 \) commutation matrix such that \( \text{vec}(B') = K_n \text{vec}(B) \) for any \( n \times n \) matrix \( B \) (Magnus & Neudecker, 2007, Ch. 3.7). Using \( \text{vec}(ABC) = (C' \otimes A) \text{vec}(B) \),
\[
\frac{\partial L_k}{\partial \text{vec}(\Psi')} = \frac{\partial L_k}{\partial \text{vec}(f_k')} \frac{\partial \text{vec}(f_k')}{\partial \text{vec}(\Psi')} = \text{vec} \left( C_k \tilde{\Psi}_k e^{i\omega k \ell} + C_k^* \overline{\Psi}_k e^{-i\omega k \ell} \right)'.
\]
Since $C_k^* = C_k$, we get $\partial L_k / \partial \Psi_\ell = 2 \text{Re} \left( C_k \tilde{\Psi}_k e^{i \omega_k \ell} \right)$, so

$$
\frac{\partial \log p_{Y | \Psi}(Y_T | \Psi)}{\partial \Psi_\ell} = -\frac{1}{2} \sum_{k=0}^{T-1} \frac{\partial L_k}{\partial \Psi_\ell}
= -\sum_{k=0}^{T-1} \text{Re} \left( C_k \sum_{\ell=1}^{q+1} e^{-i \omega_k (\ell-1)} \Psi_{\ell-1} e^{i \omega_k \ell} \right)
= -\sum_{\ell=0}^{q} \text{Re} \left( \sum_{k=0}^{T-1} C_k e^{-i \omega_k (\ell-\ell')} \right) \Psi_{\ell'}.
$$

Finally, $\sum_{k=0}^{T-1} C_k e^{-i \omega_k (\ell-\ell')} = \sum_{k=0}^{T-1} C_k e^{-i \omega_k \ell'} k = \tilde{C}_{\ell-\ell}$ for $\ell \geq \ell'$, and $\sum_{k=0}^{T-1} C_k e^{-i \omega_k (T+\ell-\ell')} = \sum_{k=0}^{T-1} C_k e^{-i \omega_k (1+\ell-\ell')} = \tilde{C}_{1-\ell}$ for $\ell < \ell'$.

**B.3 Proof of Lemma 1**

By the triangle inequality,

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

If $\hat{\Gamma} \overset{p}{\to} \Gamma_0$, 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 $U$ of $\Gamma_0$ in $\Xi$ such that $||P_{\theta | \Gamma}(\cdot | \Gamma) - P_{\theta | \Gamma}(\cdot | \Gamma_0)||_{L_1} < \epsilon/2$ for all $\Gamma \in U$. By assumption (ii), $P_{\Gamma | Y}(U^c | Y_T) < \epsilon/4$ w.p.a. 1. The decomposition (10) then implies

$$
||P_{\theta | Y}(\cdot | Y_T) - P_{\theta | \Gamma}(\cdot | \Gamma_0)||_{L_1} = \left\| \int_U \left[ P_{\theta | \Gamma}(\cdot | \Gamma) - P_{\theta | \Gamma}(\cdot | \Gamma_0) \right] P_{\Gamma | Y}(d\Gamma | Y_T) \right\|_{L_1}
\leq \int_U \left\| P_{\theta | \Gamma}(\cdot | \Gamma) - P_{\theta | \Gamma}(\cdot | \Gamma_0) \right\|_{L_1} P_{\Gamma | Y}(d\Gamma | Y_T)
+ \int_U \left\| P_{\theta | \Gamma}(\cdot | \Gamma) - P_{\theta | \Gamma}(\cdot | \Gamma_0) \right\|_{L_1} P_{\Gamma | Y}(d\Gamma | Y_T)
\leq \int_U \frac{\epsilon}{2} P_{\Gamma | Y}(d\Gamma | Y_T) + 2P_{\Gamma | Y}(U^c | Y_T)
\leq \frac{\epsilon}{2} + \frac{2\epsilon}{4}
= \epsilon
$$

w.p.a. 1. Here I use that the $L_1$ distance between probability measures is bounded by 2. □
B.4 Proof of Lemma 3

I follow the proof of Theorem 1.3.4 in Ghosh & Ramamoorthi (2003). Set $\kappa_2 = \sup_{\Gamma \in \mathcal{U}} \phi(\Gamma)$. Notice that $\hat{\phi}(\Gamma_0) = 0$ and assumption (i) together imply $\phi(\Gamma_0) = 0$. By assumptions (ii)–(iii), we can therefore find a small neighborhood $\mathcal{V}$ of $\Gamma_0$ in $\Xi_{\Gamma}$ such that $\kappa_1 = \inf_{\Gamma \in \mathcal{V}} \phi(\Gamma)$ satisfies $\max\{\kappa_2, \zeta\} < \kappa_1 < 0$. We may shrink $\mathcal{V}$ to ensure that it also satisfies $\mathcal{V} \subset \mathcal{U} \cap \mathcal{K}$. Choose $\delta > 0$ such that $\kappa_1 - \delta > \max\{\kappa_2 + \delta, \zeta\}$. Write

$$P_{\Gamma|Y}(\mathcal{U} | Y_T) = \left(1 + \frac{\int_{\mathcal{U}} e^{T\hat{\phi}(\Gamma)} \Pi_{\Gamma}(d\Gamma)}{\int_{\mathcal{U}} e^{T\phi(\Gamma)} \Pi_{\Gamma}(d\Gamma)}\right)^{-1}$$

Assumptions (i) and (iv) imply that the following three inequalities hold w.p.a. 1:

$$\sup_{\Gamma \in \mathcal{V}} \hat{\phi}(\Gamma) > \kappa_1 - \delta, \quad \sup_{\Gamma \in \mathcal{U} \cap \mathcal{K}} \hat{\phi}(\Gamma) < \kappa_2 + \delta, \quad \sup_{\Gamma \in \mathcal{K}^c} \hat{\phi}(\Gamma) < \zeta.$$

We then have

$$P_{\Gamma|Y}(\mathcal{U} | Y_T) \geq \left(1 + \frac{\int_{\mathcal{K}^c} e^{T\phi(\Gamma)} \Pi_{\Gamma}(d\Gamma) + \int_{\mathcal{U} \cap \mathcal{K}} e^{T\hat{\phi}(\Gamma)} \Pi_{\Gamma}(d\Gamma)}{\int_{\mathcal{U}} e^{T\phi(\Gamma)} \Pi_{\Gamma}(d\Gamma)}\right)^{-1}$$

w.p.a. 1. Since $\Pi_{\Gamma}(\mathcal{V}) > 0$ by assumption (v), and $\kappa_1 - \delta > \max\{\kappa_2 + \delta, \zeta\}$, I conclude that $P_{\Gamma|Y}(\mathcal{U} | Y_T) \overset{p}{\to} 1$ as $T \to \infty$.

\[ \square \]

B.5 Proof of Theorem 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.7.3, which allows me to use Lemma 4 to infer posterior consistency for $\Gamma_0$ under the Whittle likelihood.

Let $M: \mathcal{T}_{n,q} \to \mathbb{B}_{n,q} \times \mathbb{S}_n$ 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 $\Gamma(k) = \Sigma_{\ell=0}^{q-k} \beta_{\ell+k} \Sigma_{\ell} \beta_{\ell}$ (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.7.3.

I first show that the induced posterior for \((\beta_0, \Sigma_0)\) is consistent. Let \(\bar{U}\) be any neighborhood of \((\beta_0, \Sigma_0)\) in \(\mathbb{B}_{n,q} \times \mathbb{S}_n\). Since \(M(\cdot)\) is continuous, \(M^{-1}(\bar{U})\) 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 \(\Pi_{\beta,\Sigma}(\cdot)\):

\[
\Pi_{\beta,\Sigma}(\bar{U}) = \Pi_\Gamma(M^{-1}(\bar{U})) > 0.
\]

Due to Assumption 3 and the fact that \((\beta_0, \Sigma_0)\) is in the support of \(\Pi_{\beta,\Sigma}(\cdot)\), Lemma 4 implies that \(P^W_{\beta,\Sigma|Y}(\bar{U} \mid Y_T) \xrightarrow{p} 1\) for any neighborhood \(\bar{U}\) of \((\beta_0, \Sigma_0)\) in \(\mathbb{B}_{n,q} \times \mathbb{S}_n\).

I now prove posterior consistency for \(\Gamma_0\). Since \(f(\omega; M(\Gamma)) = f(\omega; \Gamma)\) for all \(\omega \in [-\pi, \pi]\) and \(\Gamma \in \mathbb{T}_{n,q}\), we have \(P^W_{\beta,\Sigma}(|Y_T \mid M(\Gamma)) = P^W_{\beta,\Sigma}(|Y_T \mid \Gamma)\) 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 \(\mathcal{U} = M(\mathcal{U})\) is a neighborhood of \((\beta_0, \Sigma_0)\) in \(\mathbb{B}_{n,q} \times \mathbb{S}_n\). It follows from Step 1 that

\[
P^W_{\gamma|Y}(\mathcal{U} \mid Y_T) = P^W_{\beta,\Sigma|Y}(\bar{U} \mid Y_T) \xrightarrow{p} 1.
\]

Moreover, the proof of Lemma 4 implies that the Whittle posterior is consistent regardless of whether the Whittle likelihood is based on integrals or discretized sums. \(\square\)

### B.6 Proof of Theorem 2

By the calculation in Eqn. 11 of Moon & Schorfheide (2012), the Whittle posterior \(P^W_{\theta|Y}(\cdot \mid Y_T)\) satisfies a decomposition of the form (10), where the posterior measure for the ACF \(\Gamma\) is given by \(P^W_{\theta|Y}(\cdot \mid Y_T)\). By Theorem 1, the latter posterior measure is consistent for \(\Gamma_0\) provided that the induced prior \(\Pi_\Gamma(\cdot)\) has \(\Gamma_0\) in its support.

\(\Gamma_0\) is indeed in the support of \(\Pi_\Gamma(\cdot)\), for the following reason. Let \(\Gamma(\Theta, \sigma)\) denote the map (7) from structural parameters \((\Theta, \sigma)\) ∈ \(\Xi_\Theta \times \Xi_\sigma\) to ACFs \(\Gamma \in \mathbb{T}_{n,q}\). There exists a (non-unique) set of IRFs and shock standard deviations \((\tilde{\Theta}, \tilde{\sigma})\) ∈ \(\Xi_\Theta \times X_\sigma\) such that \(\Gamma_0 = \Gamma(\tilde{\Theta}, \tilde{\Sigma})\) (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 \((\tilde{\Theta}, \tilde{\Sigma})\) in \(\Xi_\Theta \times \Xi_\sigma\). Because \(\Pi_{\Theta,\sigma}(\cdot)\) has
full support on $\Xi_\Theta \times \Xi_\sigma$, we have $\Pi_\Gamma(U) = \Pi_{\Theta,\sigma}(\Gamma^{-1}(U)) > 0$. Since the neighborhood $U$ was arbitrary, $\Gamma_0$ lies in the support of the induced prior $\Pi_\Gamma(\cdot)$.

Finally, note that the empirical autocovariances $\hat{\Gamma}$ are consistent for the true ACF $\Gamma_0$ under Assumption 3. Hence, the assumptions of the general Lemma 1 are satisfied for the Whittle SVMA posterior, and Theorem 2 follows.

**B.7 Proof of Lemma 4**

The proof closely follows the steps in Dunsmuir & Hannan (1976, Sec. 3) for proving consistency of the Whittle maximum likelihood estimator in a reduced-form identified VARMA model. Note that the only properties of the data generating process used in Dunsmuir & Hannan (1976, Sec. 3) are covariance stationarity and ergodicity for second moments, as in Assumption 3. Dunsmuir & Hannan also need $T^{-1}y_{t-k}y_{t-k}^T \xrightarrow{P} 0$ for fixed $t$ and $k$, which follows from Markov’s inequality under covariance stationarity. Where Dunsmuir & Hannan (1976) appeal to almost sure convergence, I substitute convergence in probability.

Define the normalized log likelihood ratio

$$
\hat{\phi}(\beta, \Sigma) = T^{-1} \log \frac{p_{\beta,\Sigma}(Y_T | \beta, \Sigma)}{p_{\beta,\Sigma}(Y_T | \beta_0, \Sigma_0)}.
$$

By the Kolmogorov-Szegö formula, for any $(\beta, \Sigma) \in \mathbb{B}_{n,q} \times \mathbb{S}_n$,

$$
\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \det(\tilde{f}(\omega; \beta, \Sigma)) \, d\omega = \log \det(\Sigma) - n \log(2\pi). \tag{17}
$$

Hence,

$$
\hat{\phi}(\beta, \Sigma) = \frac{1}{2} \log \det(\Sigma_0^{-1}) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr}\left\{[\tilde{f}(\omega; \beta_0, \Sigma_0)^{-1} - \tilde{f}(\omega; \beta, \Sigma)^{-1}]\hat{f}(\omega)\right\} \, d\omega. \tag{18}
$$

Define also the function

$$
\phi(\beta, \Sigma) = \frac{1}{2} \log \det(\Sigma_0^{-1}) + \frac{1}{2} \int_{-\pi}^{\pi} \text{tr}\{I_n - \tilde{f}(\omega; \beta, \Sigma)^{-1}\hat{f}(\omega; \beta_0, \Sigma_0)\} \, d\omega.
$$

$\phi(\beta, \Sigma)$ is continuous. By the argument in Dunsmuir & Hannan (1976, p. 5) (see also Brockwell & Davis, 1991, Prop. 10.8.1, for the univariate case), we have $\phi(\beta, \Sigma) \leq \phi(\beta_0, \Sigma_0) = 0$ for all $(\beta, \Sigma) \in \mathbb{B}_{n,q} \times \mathbb{S}_n$, with equality if and only if $(\beta, \Sigma) = (\beta_0, \Sigma_0)$.

The remainder of the proof verifies the conditions of Lemma 3 in five steps.
Step 1. I first show that there exists a neighborhood \( \mathcal{K} \) of \((\beta_0, \Sigma_0)\) in \( \mathbb{B}_{n,q} \times \mathbb{S}_n \) such that

\[
\sup_{(\beta, \Sigma) \in \mathcal{K}} |\hat{\phi}(\beta, \Sigma) - \phi(\beta, \Sigma)| = o_p(1). \tag{19}
\]

By definition of the Wold decomposition of a time series with a non-singular spectral density, all the roots of \( z \mapsto \det(\Phi(\beta_0; z)) \) lie strictly outside the unit circle. \( \tilde{f}(\omega; \beta, \Sigma)^{-1} = \Phi(\beta; e^{i\omega})^{-1}\Sigma^{-1}\Phi(\beta; e^{-i\omega})^{-1} \) is therefore uniformly continuous in \((\omega, \beta, \Sigma)\) for all \( \omega \in [-\pi, \pi] \) and \((\beta, \Sigma)\) in a small neighborhood of \((\beta_0, \Sigma_0)\). Denoting this neighborhood by \( \mathcal{K} \), the discussion around Lemma 1 in Dunsmuir & Hannan (1976, p. 350) implies (19).

Step 2. For any \((\beta, \Sigma) \in \mathbb{B}_{n,q} \times \mathbb{S}_n \) and \( z \in \mathbb{C} \), define the adjoint of \( \Phi(\beta; z) \) as

\[
\Phi_{\text{adj}}(\beta; z) = \Phi(\beta; z)^{-1} \det(\Phi(\beta; z)),
\]

so \( \tilde{f}(\omega; \beta, \Sigma) = |\det(\Phi(\beta; e^{i\omega}))|^2 \Phi_{\text{adj}}(\beta; e^{-i\omega})^{-1} \Sigma \Phi_{\text{adj}}(\beta; e^{-i\omega})^{-1*} \). The elements of \( \Phi_{\text{adj}}(\beta; z) \) are polynomials in \( z \), each polynomial of order \( \kappa \leq q(n - 1) \) (Dunsmuir & Hannan, 1976, p. 354). Write the matrix polynomial as \( \Phi_{\text{adj}}(\beta; z) = I_n + \sum_{\ell=1}^{\kappa} \beta_{\text{adj},\ell} z^\ell \), and define \( \tilde{\Phi}_{\text{adj}}(\beta) = (\sum_{\ell=1}^{\kappa} \|\beta_{\text{adj},\ell}\|^2)^{1/2} \).

Now define, for \( \delta \geq 0 \),

\[
\tilde{f}_\delta(\omega; \beta, \Sigma) = (|\det(\Phi(\beta; e^{-i\omega}))|^2 + \delta) \Phi_{\text{adj}}(\beta; e^{-i\omega})^{-1} \Sigma \Phi_{\text{adj}}(\beta; e^{-i\omega})^{-1*},
\]

\[
\hat{\phi}_\delta(\beta, \Sigma) = \frac{1}{2} \log \det(\Sigma_0 \Sigma^{-1}) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ [\tilde{f}(\omega; \beta_0, \Sigma_0)^{-1} - \tilde{f}_\delta(\omega; \beta, \Sigma)^{-1}] \hat{I}(\omega) \right\} d\omega,
\]

and

\[
\phi_\delta(\beta, \Sigma) = \frac{1}{2} \log \det(\Sigma_0 \Sigma^{-1}) + \frac{1}{2} \int_{-\pi}^{\pi} \text{tr} \{ I_n - \tilde{f}_\delta(\omega; \beta, \Sigma)^{-1} \tilde{f}(\omega; \beta_0, \Sigma_0) \} d\omega.
\]

Because \( \hat{I}(\omega) \) is positive semidefinite for each \( \omega \in [-\pi, \pi] \), we have \( \hat{\phi}(\beta, \Sigma) \leq \hat{\phi}_\delta(\beta, \Sigma) \) for all \((\beta, \Sigma) \in \mathbb{B}_{n,q} \times \mathbb{S}_n \) and \( \delta > 0 \).

Finally, for any \( c_1, c_2, c_3 > 0 \), define the set

\[
\tilde{\mathcal{K}}(c_1, c_2, c_3) = \{ (\beta, \Sigma) \in \mathbb{B}_{n,q} \times \mathbb{S}_n : \lambda_{\text{min}}(\Sigma) \geq c_1, \|\Sigma\| \leq c_2, \tilde{\Phi}_{\text{adj}}(\beta) \leq c_3 \},
\]

where \( \lambda_{\text{min}}(\Sigma) \) is the smallest eigenvalue of \( \Sigma \).
The discussion surrounding Lemma 3 in Dunsmuir & Hannan (1976, p. 351) then gives
\[
\sup_{(\beta, \Sigma) \in K(c_1, c_2, c_3)} |\hat{\phi}_3(\beta, \Sigma) - \phi_3(\beta, \Sigma)| = o_p(1),
\]
for any \(c_1, c_2, c_3 > 0\). Because \(\phi_3(\beta, \Sigma)\) is continuous in \((\beta, \Sigma, \delta)\) at \((\beta = \beta_0, \Sigma = \Sigma_0, \delta = 0)\), and \(\phi(\beta, \Sigma) = \phi_{\delta=0}(\beta, \Sigma)\) is uniquely maximized at \((\beta_0, \Sigma_0)\),
\[
\inf_{\delta > 0} \sup_{(\beta, \Sigma) \notin K} \phi_3(\beta, \Sigma) < \phi(\beta_0, \Sigma_0) = 0.
\]
I conclude that for all \(c_1, c_2, c_3 > 0\) there exist \(\delta > 0\) and \(\zeta > 0\) such that
\[
\sup_{(\beta, \Sigma) \in K(c_1, c_2, c_3) \cap K^c} \hat{\phi}(\beta, \Sigma) \leq \sup_{(\beta, \Sigma) \in K(c_1, c_2, c_3) \cap K^c} \phi_3(\beta, \Sigma) \leq -\zeta + o_p(1).
\]

**Step 3.** Let \(\zeta > 0\) be the scalar found in the previous step. The proof of Theorem 4(i) in Dunsmuir & Hannan (1976, pp. 354–355) (see also the beginning of the proof of their Theorem 3, pp. 352–353) shows that there exist \(c_1, c_2, c_3 > 0\) such that
\[
\sup_{(\beta, \Sigma) \in K(c_1, c_2, c_3) \cap K^c} \hat{\phi}(\beta, \Sigma) \leq -\zeta.
\]

**Step 4.** Steps 1–3 imply that the sufficient conditions in Lemma 3 hold. I conclude that \(P_{W,\Sigma Y}(\tilde{U} | Y_T) \xrightarrow{p} 1\) for any neighborhood \(\tilde{U}\) of \((\beta_0, \Sigma_0)\) in \(\mathbb{R}_{+} \times \mathbb{S}_n\).

**Step 5.** Finally, I prove an assertion in Appendix A.7.3: Lemma 4 holds for the discretized Whittle likelihood that replaces integrals \((2\pi)^{-1} \int_0^{2\pi} g(\omega) d\omega\) (for a \(2\pi\)-periodic function \(g(\cdot)\)) in the definition of \(\log p_{W,\Sigma Y}^Y(Y_T | \beta, \Sigma)\) with sums \(T^{-1} \sum_{k=0}^{T-1} g(\omega_k), \omega_k = 2\pi k/T\).

The proof of Theorem 4(ii) of Dunsmuir & Hannan (1976, p. 356) shows that steps 1–3 above carry through if the integral in expression (18) is replaced with a discretized sum. The only other effect of discretizing the integrals in the Whittle likelihood is that the Kolmogorov-Szegö formula (17) does not hold exactly. Instead,
\[
T^{-1} \sum_{j=0}^{T-1} \log \det(\hat{f}(\omega_j; \beta, \Sigma)) = \log \det(\Sigma) - n \log(2\pi) + T^{-1} \sum_{j=0}^{T-1} \log |\det(\Phi(\beta; e^{-i\omega_j}))|^2.
\]

The posterior consistency result for the discretized Whittle posterior follows from steps 1–4
above if I show
\[
\sum_{j=0}^{T-1} \log |\det(\Phi(\beta; e^{-i\omega_j}))|^2 \leq 2nq \log 2 \tag{20}
\]
for all \((\beta, \Sigma) \in \mathbb{B}_{n,q} \times \mathbb{S}_n\), and furthermore,
\[
\sum_{j=0}^{T-1} \log |\det(\Phi(\beta; e^{-i\omega_j}))|^2 = o_p(1) \tag{21}
\]
uniformly in a small neighborhood of \((\beta_0, \Sigma_0)\) in \(\mathbb{B}_{n,q} \times \mathbb{S}_n\).

For any \(\beta \in \mathbb{B}_{n,q}\) and \(z \in \mathbb{C}\), write \(\det(\Phi(z; \beta)) = \det(I_n + \sum_{\ell=1}^q \beta \ell z^\ell) = \prod_{b=1}^{nq} (1 - a_b(\beta)z)\) for some complex scalars \(\{a_b(\beta)\}_{1 \leq b \leq nq}\) that depend on \(\beta\) and satisfy \(|a_b(\beta)| < 1\) (Brockwell & Davis, 1991, p. 191). From the Taylor series \(\log(1 - z) = -\sum_{s=1}^{\infty} z^s/s\) (valid for \(z \in \mathbb{C}\) inside the unit circle) we get, for all \(\beta \in \mathbb{B}_{n,q}\),
\[
\sum_{k=0}^{T-1} \log \det(\Phi(e^{-i\omega_k}; \beta)) = -\sum_{k=0}^{T-1} \sum_{b=1}^{nq} \sum_{s=1}^{\infty} \frac{(a_b(\beta)e^{-i\omega_k})^s}{s} = -\sum_{b=1}^{nq} \sum_{s=1}^{\infty} \frac{(a_b(\beta))^s}{s} \sum_{k=0}^{T-1} e^{-i\omega_k s}.
\]
Since \(\sum_{k=0}^{T-1} e^{-i\omega_k s}\) equals \(T\) when \(s\) is an integer multiple of \(T\), and equals 0 otherwise,
\[
\sum_{k=0}^{T-1} \log \det(\Phi(e^{-i\omega_k}; \beta)) = -\sum_{b=1}^{nq} \sum_{s=1}^{\infty} \frac{(a_b(\beta))^s}{s} = \sum_{b=1}^{nq} \log \left(1 - (a_b(\beta))^T\right).
\]
Hence,
\[
\sum_{k=0}^{T-1} \log |\det(\Phi(e^{-i\omega_k}; \beta))|^2 = \sum_{k=0}^{T-1} \log |\det(\Phi(e^{-i\omega_k}; \beta))|^2 + \sum_{k=0}^{T-1} \log |\det(\Phi(e^{-i\omega_k}; \beta)^*)| = \sum_{b=1}^{nq} \log |1 - (a_b(\beta))^T|^2 \leq nq \log 4,
\]
where the inequality uses \(|1 - (a_b(\beta))^T| < 2\). Claim (20) follows. For \(\beta\) in a small neighborhood of \(\beta_0\), \(\max_b |a_b(\beta)|\) is uniformly bounded away from 1. This implies claim (21). \(\square\)
References


Journal of Econometrics 64, 183–206.


