Title

Dynamic Causal Inference with Imperfect Identifying Information

Author

Nguyen, Lam Hoang

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Dynamic Causal Inference with Imperfect Identifying Information

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Lam Nguyen

Committee in charge:

Professor James D. Hamilton, Chair
Professor Brendan Beare
Professor Graham Elliott
Professor Allan Timmermann
Professor Rossen Valkanov
Professor Ying Zhu

2020

The dissertation of Lam Nguyen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my adviser, James D. Hamilton, and my econometric professors, Brendan Beare, and Graham Elliott. In each of their own ways, they have taught me how to think deeply about research questions and examine the evidence critically. Their unwavering support has been indispensable to me in times of difficulties and will always inspire me to aim higher. I am also grateful to Allan Timmermann, Rossen Valkanov, Ying Zhu, Valerie Ramey, David Lagakos, Marjorie Flavin, Johannes Wieland, Yixiao Sun, and Kaspar Wuthrich, who have helped and advised me in various ways.

On a more personal level, I would also like to thank many friends whom I met in San Diego including but not limited to Qihui Chen, Wenbin Wu, Yanjun Liao, Xu Zhang, Alex Yu, Daniel Leff-Yaffe, Julian Martinez-Iriarte, Alejandro Nakab, Bruno Lopez-Videla Mostajo, Sameem Siddiqui and Alex Kellog. I am most fortunate to have their company in what is known to be a lonely journey. I am also grateful to many other teachers and colleagues who have influenced me over the years. In particular, I am deeply indebted to James Morsink, Prachi Mishra, James Miles, Sung Jae Jun, and Herman Bierens, who led me into the field of economics and econometrics.

Most important, I am grateful to my parents and my sister for always encouraging me to pursue what I want, and for the enormous sacrifices they made for me.

Chapter 1, in full, is currently being prepared for submission for publication of the material. Lam Nguyen. The dissertation author was the primary author of this chapter.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Lam Nguyen. The dissertation author was a primary author of this chapter.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Lam Nguyen. The dissertation author was a primary author of this chapter.

VITA

2012	B. S. in Finance, *(with Highest Distinction)*, Pennsylvania State University
2012	Minor in Economics, Pennsylvania State University
2015-2020	Teaching Assistant, University of California San Diego
2020	Ph. D. in Economics, University of California San Diego

FIELDS OF STUDY

Major Field: Econometrics, Macroeconomics

ABSTRACT OF THE DISSERTATION

Dynamic Causal Inference with Imperfect Identifying Information

by

Lam Nguyen

Doctor of Philosophy in Economics

University of California San Diego, 2020

Professor James D. Hamilton, Chair

This dissertation contains three essays exploring how macroeconomists can identify and estimate dynamic causal effects in models where researchers have doubts about identifying assumptions.

Chapter 1 proposes a new Markov Chain Monte Carlo algorithm to estimate a sign-restricted structural vector autoregression on time series that are subject to regime shifts. My approach can incorporate useful prior information about both model parameters and hidden states while transparently imposing sign restrictions. I illustrate my method by revisiting the literature on asymmetric effects of conventional monetary policy during recessions and expansions. My evidence suggests that previous empirical research found asymmetric effects by questionable

identification schemes and neglecting changes in the variances of structural shocks. I find little difference in the structural parameters, and thus I do not find evidence of asymmetry.

Chapter 2 studies the method of instrumental variables in set-identified models. I develop a proxy structural vector autoregression in which prior information from both theory and the empirical literature is incorporated about signs and magnitudes of certain parameters and equilibrium impacts. I use my method to investigate the relevance and validity of three popular instruments for monetary policy shocks, developed by Romer and Romer (2004), Sims and Zha (2006), and Smets and Wouters (2007). I find that all of them are strongly relevant but only that of Smets and Wouters is valid. Furthermore, the empirical analysis demonstrates that my framework can combine information from a relevant and valid instrument with prior information about sign restrictions to improve inference about structural impulse-response functions.

Chapter 3 develops new methods to study dynamic causal effects in a data-rich environment. Current development in high-dimensional statistics fails to address the main interest of economists: causal inference with credible assumptions. I first review the literature on high-dimensional linear regression models and dynamic factor models. Then, I develop several new Bayesian numerical algorithms that combine the techniques in high-dimensional statistics with recent advances in dynamic causal inference. In particular, I discuss how to make causal statements from a high-dimensional structural model when researchers have doubts about identifying assumptions. Finally, I extend those algorithms to the case of Markov-switching models to accommodate nonlinearities in economic time series.

Chapter 1

Regime-Switching Structural Vector Autoregression Identified by Sign Restrictions: Asymmetric Effects of Monetary Policy Revisited

Abstract

Economic relations change over time and with the business cycle. This paper proposes a new Markov Chain Monte Carlo algorithm to estimate a sign-restricted structural vector autoregression on time series that are subject to regime shifts. My approach can incorporate useful prior information about both model parameters and hidden states while transparently imposing sign restrictions. I illustrate my method by revisiting the literature on asymmetric effects of conventional monetary policy during recessions and expansions. My evidence suggests that previous empirical research found asymmetric effects by questionable identification schemes and neglecting changes in the variances of structural shocks. I find little difference in the structural parameters, and thus I do not find evidence of asymmetry.

1.1 Introduction

Empirical macroeconomists are interested in estimating *dynamic causal effects* of unobserved structural shocks on endogenous macroeconomic variables.[1] The Vector Autoregression (VAR) has been one of the main tools to accomplish this task. Since the pioneering paper by Sims (1980), this research agenda has evolved to deal with two main challenges: (1) the *identification problem* and (2) *nonlinearity* of macroeconomic time series.

Causal inference in VAR always requires *identifying assumptions* because there are multiple structural models consistent with the observed data. As documented in Ramey (2016) and Nakamura & Steinsson (2018), conventional exclusion restrictions are often both conceptually and empirically controversial, and consequently, economists have been searching for new identification strategies that only impose minimal, non-controversial identifying assumptions. A new promising strategy is *identification by sign restrictions* where researchers only assert their beliefs about the signs, instead of the magnitudes, of various quantities. This literature is pioneered by Uhlig (2005), Faust (1998), and Canova & De Nicolo (2002), and since then it has been rising with many notable contributions in both theories and applications, including Rubio-Ramírez, Waggoner & Zha (2010), Arias, Rubio-Ramírez & Waggoner (2018b), Arias, Caldara & Rubio-Ramírez (2019), Antolín-Díaz & Rubio-Ramírez (2018), and Baumeister & Hamilton (2015), Baumeister & Hamilton (2018), and Baumeister & Hamilton (2019).

Nevertheless, this recent surge in the literature has mainly focused on linear VAR. Ignoring potential nonlinearity is an important omission because most macroeconomic time series exhibit dramatic breaks (Hamilton (1989) and Hamilton (2016)). Primiceri (2005), Cogley & Sargent (2001) and Cogley & Sargent (2005), and Sims & Zha (2006) all found that time-varying parameter VARs are better descriptions of the economy than their linear counterparts. The literature on time-varying parameter models is also vast, and there is considerable interest and promise in

[1] I adopt Ramey (2016)'s definition of structural shocks. Specifically, structural shocks must reflect either current or future unanticipated movement of exogenous variables and be independent of both endogenous variables and each other.

modeling the time variation as changes in regime (Owyang & Ramey (2004), Benati & Surico (2009), Lanne, Lütkepohl & Maciejowska (2010), Herwartz & Lütkepohl (2014)). With few exceptions, most papers still use exclusion restrictions for identification and hence their results are fragile. Current applications of identification by sign restriction to time-varying parameter VAR are limited and suffer from two shortcomings: (1) inference is sensitive to the ordering of variables (Baumeister & Peersman (2013)), and (2) the algorithm imposes implicit priors on the dynamic causal effects (Rubio-Ramírez, Waggoner & Zha (2005), Bognanni (2018)).

This paper extends Baumeister & Hamilton (2015)'s SVAR framework to identify and estimate *regime-dependent dynamic causal effects*. Specifically, I develop a new Markov Chain Monte Carlo (MCMC) method to estimate an *exogenous Markov-switching structural VAR (SVAR) model* identified by *sign restrictions*. By allowing the parameters to evolve according to a Markov chain, this framework captures any potential regime-dependent effects, and the use of sign restrictions ensures that its identification strategy is more robust than traditional approaches. Compared to Sims & Zha (2006), my algorithm can implement not only conventional short-run and long-run restrictions but also sign restrictions. And in contrast to Rubio-Ramírez, Waggoner & Zha (2005) and Bognanni (2018), my procedure transparently incorporates priors and sign restrictions without using draws from the uniform Haar distribution to rotate the reduced-form VAR, a practice known to impose implicit priors on the dynamic causal effects.[2] Moreover, borrowing insights from Frühwirth-Schnatter (2001), my method explicitly deals with the *label-switching problem* in Bayesian estimation of mixture models. This well-known statistical problem results in slow convergence of the MCMC sampler and biased inference but has been ignored in the previous econometric literature (Stephens (2000), Celeux, Hurn & Robert (2000)).

Furthermore, my algorithm can improve inference about the hidden states by incorporating qualitative information constructed from external sources, such as recession indicators from the National Bureau of Economic Research (NBER) or monetary policy stance indicators from

[2]See Baumeister & Hamilton (2015) and Baumeister & Hamilton (2018) for detailed discussions.

Boschen & Mills (1995). Empirical work often uses external information dogmatically by assuming that the qualitative indicators are perfectly accurate.[3] My Bayesian approach generalizes this practice by allowing researchers to express their uncertainty about the quality of those indicators. This technique is introduced by Jefferson (1998) and recently used in Lyu & Noh (2018). The difference between those studies and mine is that they implement it in the frequentist framework whereas mine is Bayesian. Besides allowing researchers to incorporate their priors about the quality of the indicators, my framework can formally test the usefulness of those indicators by Bayesian model comparison.

To demonstrate my method, I revisit the literature on asymmetric effects of conventional monetary policy during recessions and expansions.[4] This question has received renewed interest among economists and policy makers in the aftermath of the Global Financial Crisis (Mishkin (2009)).[5] Despite the vast literature, no consensus has emerged. Theoretical models that result in a convex aggregate supply curve (Ball & Mankiw (1994)), credit market imperfection (Bernanke, Gertler & Gilchrist (1999)), or loss aversion (Santoro, Petrella, Pfajfar & Gaffeo (2014)) imply that monetary policy has stronger effects in recessions, whereas recent theories that incorporate firm level idiosyncratic volatility (Vavra (2013)) suggest the opposite. Empirical evidence from Kaufmann (2002), Garcia & Schaller (2002), Lo & Piger (2005), and Peersman & Smets (2002) and Peersman & Smets (2005) suggests that monetary policy has stronger effects in recessions, whereas Thoma (1994) and Tenreyro & Thwaites (2016) conclude it does not.

I analyze the asymmetric effects of monetary policy by estimating a 2-state Markov-switching SVAR model of the output gap, inflation, and Fed fund rate over the period from 1950 to 2007. My method has two main improvements over previous work. First, I use the sign

[3] For example, Owyang, Ramey & Zubairy (2013) and Ramey & Zubairy (2018) determine recessions as periods when the unemployment rate is higher than 6.5%.

[4] In this paper, I don't consider the asymmetric effects with respect to the *size* or *sign* of monetary policy shocks, such as those in Cover (1992), Karras (1996), Barnichon, Matthes & Ziegenbein (2016), and Angrist, Jordà & Kuersteiner (2018)

[5] There is a related literature on asymmetric effects of fiscal policy during recessions and expansions, including Auerbach & Gorodnichenko (2012a), Auerbach & Gorodnichenko (2012b), and Auerbach & Gorodnichenko (2013), Owyang, Ramey & Zubairy (2013), Ramey & Zubairy (2018), and Lyu & Noh (2018).

restrictions developed by Baumeister & Hamilton (2018) to identify structural shocks whereas earlier work typically uses changes in endogenous variables, innovations from a reduced-form model, or shocks identified by exclusion restrictions. Thus, my identification strategy is more robust and consistent with a wider range of structural models. Second, my model allows for a more general form of regime changes, and it distinguishes regime changes in structural parameters from those in the variances of structural shocks. This distinction is especially important because failure to incorporate heteroskedasticity will exaggerate changes in coefficients (Stock (2001) and Sims & Zha (2006)). Indeed, most previous work only allows the coefficients of interest to change while ignoring potential heteroskedasticity in the innovations. As a result, their analyses incorrectly favor asymmetry.

The empirical analysis reveals that the regime-switching model is a better description of the data than its linear counterpart. However, the major difference lies in the variances of the structural shocks during recessions and expansions. My results show that structural shocks have larger variances in recessions, implying that neglecting these changes will falsely magnify the effects of monetary policy during these periods. Indeed, I find symmetric effects of monetary policy across both regimes. In particular, monetary tightening reduces the output gap and inflation by similar magnitudes in both recessions and expansions. My results remain robust even when I augment the baseline model with the NBER recession indicator and use asymmetric priors for structural parameters.

The rest of this paper is organized as follows. Section 1.2 formally describes the framework and the MCMC algorithm. Section 1.3 applies this method to investigate the effects of monetary policy during recessions and expansions. Section 1.4 briefly concludes. Additional technical details are shown in the appendices.

1.2 Bayesian estimation of regime-switching structural vector autoregression with sign restrictions

This section describes the regime-switching SVAR model with sign restrictions together with the estimation procedure. My MCMC algorithm incorporates insights from the Gibbs sampler for Markov-switching SVAR of Hamilton (2016) and Kim & Nelson (1999), the MCMC algorithm for sign-restricted SVAR of Baumeister & Hamilton (2015), and the permutation sampler of Frühwirth-Schnatter (2006).

1.2.1 General formulation

Suppose that the dynamics of the data can be summarized by an S-state p-th order Markov-switching SVAR

$$\mathbf{A}_{S_t}\mathbf{y}_t = \mathbf{k}_{S_t} + \mathbf{B}_{1,S_t}\mathbf{y}_{t-1} + \mathbf{B}_{2,S_t}\mathbf{y}_{t-2} + \cdots + \mathbf{B}_{p,S_t}\mathbf{y}_{t-p} + \mathbf{u}_t$$

$$\mathbf{u}_t | S_t \sim N\left(\mathbf{0}, \mathbf{D}_{S_t}\right)$$

$$P\left(S_1 = s\right) = \mu_s$$

$$P\left(S_t = s' | S_{t-1} = s\right) = p_{s',s}$$

where $S_t \in \{1, 2, \ldots, S\}$ is the indicator for the regime, $\mathbf{y}_t$ is an $(n \times 1)$ vector of observed variables, and $\mathbf{u}_t$ is a $(n \times 1)$ vector of structural shocks at time t. Structural shocks should be mutually uncorrelated, thus $\mathbf{D}_{S_t}$ is assumed to be an $(n \times n)$ diagonal matrix. When $S_t = s$, $\mathbf{A}_s$ is an $(n \times n)$ matrix that governs contemporaneous relationship between observed variables in state s, $\mathbf{k}_s$ and $\mathbf{B}_{j,s}$ $(j = 1, \ldots, p)$ are the $(n \times 1)$ vector of constants and $(n \times n)$ matrices of lag coefficients in state s, and $\mathbf{D}_s$ is an $(n \times n)$ matrix that represents the variance of the structural shocks in state s. For simplicity, the model is rewritten in a more compact form as

$$\mathbf{A}_{S_t}\mathbf{y}_t = \mathbf{B}_{S_t}\mathbf{x}_{t-1} + \mathbf{u}_t \tag{1.1}$$

where $\mathbf{B}_{S_t} \equiv \begin{bmatrix} \mathbf{B}_{1,S_t} & \dots & \mathbf{B}_{p,S_t} & \mathbf{k}_{S_t} \end{bmatrix}$ is $[n \times (np+1)]$ matrix, and $\mathbf{x}_{t-1} = \begin{bmatrix} \mathbf{y}'_{t-1} & \dots & \mathbf{y}'_{t-p} & 1 \end{bmatrix}'$ is $[(np+1) \times 1]$ vector.

We may also have K independent indicators for the hidden regimes, $\left\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(K)}\right\}$. The quality of information in the k^{th} indicator is summarized in the following $S \times S$ matrix

$$\mathbf{G}^{(k)} = \begin{bmatrix} P\left(Z_t^{(k)} = 1 | S_t = 1\right) & \dots & P\left(Z_t^{(k)} = 1 | S_t = S\right) \\ \vdots & \ddots & \vdots \\ P\left(Z_t^{(k)} = S | S_t = 1\right) & \dots & P\left(Z_t^{(k)} = S | S_t = S\right) \end{bmatrix} \tag{1.2}$$

where the s^{th} column of $\mathbf{G}^{(k)}$ summarizes the quality of information in $\mathbf{z}^{(k)}$ about regime s, thus each column sums to unity. We can express our confidence about the quality of information in $\mathbf{z}^{(k)}$ as priors on elements of $\mathbf{G}^{(k)}$. There are two important special cases: (1) Perfect indicator: $g_{s,s}^{(k)} = 1$ for all $s = 1, \dots, S$, and (2) Uninformative indicator: $\mathbf{G}_{\cdot,s}^{(k)} = \mathbf{G}_{\cdot,s'}^{(k)}$ for all $s, s' = 1, \dots, S$.

1.2.2 Discussion of econometric challenges

In SVAR analysis, dynamic causal effects are represented by *structural impulse-response functions* (SIRFs), which are functions of the model parameters. Hence, to draw causal inference from a regime-switching SVAR model, we must answer three questions: (1) What identifying assumptions do we need in order to identify the parameters and the SIRFs in a given regime?, (2) What identifying assumptions do we need in order to distinguish parameters of one regime from those of another?, and (3) Given these identifying assumptions, how do we estimate the model parameters and the hidden regimes? The first two questions require economic knowledge to answer, whereas the last question is purely a statistical problem.

First, how do we identify model parameters in a given regime? The challenge is that even without regime changes, the model is still *under-identified* because the number of unknown parameters is usually greater than the number of moment conditions. This is the *identification problem* in the SVAR literature. The conventional approach is to put quantitative restrictions on the unknown parameters to achieve point-identification, however, those restrictions might be controversial. My main innovation here is to replace those quantitative restrictions with qualitative ones. My framework can not only impose conventional identifying assumptions such as short-run and long-run restrictions, but it can also impose sign restrictions on both the elements of $\mathbf{A}$ and $\mathbf{A}^{-1}$. For example, sign restrictions on the systematic component of monetary policy in Baumeister & Hamilton (2018) and Arias, Caldara & Rubio-Ramírez (2019) are restrictions on $\mathbf{A}$, whereas sign restrictions on the contemporaneous impacts of monetary policy shock in Baumeister & Hamilton (2018) are restrictions on $\mathbf{A}^{-1}$.

Although those restrictions are more likely to be agreed by most economists, we no longer have point-identification but only *set-identification*, meaning there will be many models consistent with both the observed data and the sign restrictions. Estimation and inference in set-identified model is challenging from the frequentist perspective, and hence this paper opts for the Bayesian approach.[6] The advantage is that we can still use Bayes' rule to derive the posterior distributions as long as we specify the priors over all unknown parameters. However, there is no Bayesian free-lunch. The caveats are that some quantities of the posterior distribution will never be updated and Bayesian credible set typically no longer satisfy frequentist coverage even with an infinite amount of data. The first insight dates back to Lindley (1957) and Poirier (1998), while the latter one is illustrated in Moon & Schorfheide (2012). To overcome this setback, I propose to openly acknowledge that our posteriors will always be influenced by our priors as advocated by Baumeister & Hamilton (2015) and recommend using *informative priors* that are carefully constructed from theories and the previous empirical literature.

Second, how do we distinguish the parameters of one regime from those of another? The

[6]For discussions of frequentist inference in sign-restricted SVAR, readers are referred to Gafarov, Meier & Olea (2018), Gafarov, Meier & Olea (2016), and Granziera, Moon & Schorfheide (2018).

issue here is that the likelihood function is symmetric with respect to permutation of the parameters between the states. If one researcher comes up with some particular estimates, another researcher can swap these estimates across states and come up with another answer that is equally consistent with the observed data. This is the *normalization problem* in econometrics as investigated in Hamilton, Waggoner & Zha (2007). Their insight is that the choice of normalization will have nontrivial influence in both frequentist and Bayesian inferences. A good normalization rule should restrict the parameter space into a subset where the likelihood function/posterior distribution is well-behaved (e.g. unimodal). However, a good rule is hard to find, especially when the parameter space is large. I propose to resolve this normalization issue by selecting a rule based on our economic knowledge. For example, if we believe the hidden states to correspond to economic recessions and expansions, we can distinguish regimes by the variances of their structural shocks. Another complementary solution is to use exploratory analysis of posterior draws as in Frühwirth-Schnatter (2006).

Finally, given those identifying assumptions, how do we estimate the model? The Bayesian objective in estimation is to sample from the posterior distributions, which reflect our beliefs about the model parameters after observing the data. Here the challenge is that we only observe the endogenous variables, but we do not see the hidden regime $\mathbf{S}_T$. Because the posterior distributions cannot be characterized analytically, direct sampling cannot be used. To resolve this problem, I use MCMC methods. The general idea is to replace direct sampling from an intractable posterior distribution with iterative sampling from a sequence of ergodic Markov chain whose stationary distribution is the true posterior distribution.

How do we construct such a sequence? The key insights here are *conditioning* and *data augmentation*. Conditioning on the hidden states, the transition probabilities are point-identified and the model is linear. This observation suggests that the hidden states should be treated as unknown parameters and sampled jointly with all the others; this is the essence of data augmentation introduced by Tanner & Wong (1987). The MCMC algorithm that takes full advantage of the above insights is the *Gibbs sampler*, pioneered by Geman & Geman (1987) and popularized by Gelfand & Smith (1990). Its main insight is that iterative draws from *conditional posterior distributions* will indeed form an ergodic Markov chain whose stationary distribution is the true *unconditional posterior distribution*.

Another subtle complication with Bayesian estimation of Markov-switching model is the *label-switching problem*: since the posterior distributions are necessarily multimodal, the sampler might jump between different modes in an unbalanced fashion, resulting in extremely slow convergence. To deal with this problem, I use the *permutation sampler* of Frühwirth-Schnatter (2006). The main idea is to first sample from the *unconstrained posterior* by random permutations at the end of each Gibbs iteration, thus replacing unbalanced label switching with a balanced one. Then, a suitable normalization rule will be applied to the posterior draws to identify the subspace of interest.[7]

1.2.3 MCMC algorithm for model estimation

Denote the data for both the endogenous variables and the regime indicators as $\mathbf{W}_T = \left(\mathbf{Y}_T, \mathbf{Z}_T^{(1)}, \ldots, \mathbf{Z}_T^{(K)}\right)$ where $\mathbf{Y}_T = \{\mathbf{y}_t\}_{t=1}^T$ and $\mathbf{Z}_T^{(k)} = \left\{\mathbf{z}_t^{(k)}\right\}_{t=1}^T$. Let $\mathbf{S}_T = \{S_t\}_{t=1}^T$ be the vector of the hidden states, $\mathbf{G}$ be the vector that collects all parameters of the regime indicators, $\boldsymbol{\mu}$ be the vector of initial probabilities, $\mathbf{P} = \left\{p_{s,s'}\right\}$ be the matrix of transition probabilities, and $\boldsymbol{\Phi}$ be the vector of structural parameters across all states.[8] My MCMC algorithm is formally stated below.

MCMC Algorithm. We start the algorithm with random initial values. Suppose we are at the m^{th} iteration with parameters $\left(\mathbf{G}^{(m)}, \boldsymbol{\rho}^{(m)}, \mathbf{P}^{(m)}, \boldsymbol{\Phi}^{(m)}, (\mathbf{S}_T)^{(m)}\right)$ and we want to draw $\left(\mathbf{G}^{(m+1)}, \boldsymbol{\rho}^{(m+1)}, \mathbf{P}^{(m+1)}, \boldsymbol{\Phi}^{(m+1)}, (\mathbf{S}_T)^{(m+1)}\right)$. We do so using the following steps

[7] Geweke (2007), Celeux, Hurn & Robert (2000), and Jasra, Holmes & Stephens (2005) provide different perspective and methods to deal with the label-switching problems.

[8] $\mathbf{G} = \left[\text{vec}\left(\mathbf{G}^{(1)}\right)' \quad \text{vec}\left(\mathbf{G}^{(2)}\right)' \quad \ldots \quad \text{vec}\left(\mathbf{G}^{(k)}\right)'\right]'$.

$\mathbf{P}$ is the $S \times S$ matrix whose row s' column s element is $p_{s',s} = P(S_{t+1} = s' | S_t = s)$, thus each column sums to unity.

$\boldsymbol{\Phi} = \left[\text{vec}(\mathbf{A}_1)' \quad \text{vec}(\mathbf{B}_1)' \quad \text{vec}(\text{diag}(\mathbf{D}_1))' \quad \ldots \quad \text{vec}(\mathbf{A}_S)' \quad \text{vec}(\mathbf{B}_S)' \quad \text{vec}(\text{diag}(\mathbf{D}_S))'\right]'$

1. **Step 1**: Draw the parameters for the quality of regime indicators, $\mathbf{G}^{(m+1)}$, from their conditional posterior distribution

   $\mathbf{G}|\boldsymbol{\mu}^{(m)}, \mathbf{P}^{(m)}, \boldsymbol{\Phi}^{(m)}, (\mathbf{S}_T)^{(m)}, \mathbf{W}_T$.

   This step will be skipped if we don't use any regime indicator.

2. **Step 2**: Draw the initial probabilities, $\boldsymbol{\mu}^{(m+1)}$, from their conditional posterior distribution

   $\boldsymbol{\mu}|\mathbf{G}^{(m+1)}, \mathbf{P}^{(m)}, \boldsymbol{\Phi}^{(m)}, (\mathbf{S}_T)^{(m)}, \mathbf{W}_T$

   Draw the transition probabilities, $\mathbf{P}^{(m+1)}$, from their conditional posterior distribution

   $\mathbf{P}|\mathbf{G}^{(m+1)}, \boldsymbol{\mu}^{(m+1)}, \boldsymbol{\Phi}^{(m)}, (\mathbf{S}_T)^{(m)}, \mathbf{W}_T$

3. **Step 3**: Draw the structural parameters, $\boldsymbol{\Phi}^{(m+1)}$, from their conditional posterior distribution

   $\boldsymbol{\Phi}|\mathbf{G}^{(m+1)}, \boldsymbol{\mu}^{(m+1)}, \mathbf{P}^{(m+1)}, (\mathbf{S}_T)^{(m)}, \mathbf{W}_T$

4. **Step 4**: Draw the hidden states,$(\mathbf{S}_T)^{(m+1)}$, from their conditional posterior distribution

   $\mathbf{S}_T|\mathbf{G}^{(m+1)}, \boldsymbol{\mu}^{(m+1)}, \mathbf{P}^{(m+1)}, \boldsymbol{\Phi}^{(m+1)}, \mathbf{W}_T$

5. **Step 5**: Randomly permute all parameters and hidden states using the permutation sampler.

The above cycle is repeated a large number of times to ensure convergence, then half the simulated draws are discarded to remove the effect of the initial conditions. The remaining half will be thought of as draws from the true posterior distribution and used for subsequent analysis.

To further simplify each step of the algorithm, I assume (a) independent priors across $\mathbf{G}$, $\boldsymbol{\mu}$, $\mathbf{P}$, and $\boldsymbol{\Phi}$; and (b) the regime indicators are conditionally independent from the endogenous variables. With those two assumptions, the joint density of the data, parameters, and regimes can be decomposed as

$$P(\mathbf{W}_T, \mathbf{G}, \boldsymbol{\mu}, \mathbf{P}, \boldsymbol{\Phi}, \mathbf{S}_T) = \left[P(\mathbf{Y}_T, \mathbf{Z}_T | \mathbf{G}, \boldsymbol{\mu}, \mathbf{P}, \boldsymbol{\Phi}, \mathbf{S}_T)\right]\left[P(\mathbf{G}) P(\boldsymbol{\mu}, \mathbf{P}) P(\mathbf{S}_T | \boldsymbol{\mu}, \mathbf{P}) P(\boldsymbol{\Phi})\right] \quad (1.3)$$

$$= \underbrace{\left[\Pi_{k=1}^{K} P\left(\mathbf{Z}_T^{(k)} | \mathbf{G}^{(k)}, \mathbf{S}_T\right) P\left(\mathbf{G}^{(k)}\right)\right]}_{\text{Indicator parameters}} \underbrace{\left[P(\boldsymbol{\mu}, \mathbf{P}) P(\mathbf{S}_T | \boldsymbol{\mu}, \mathbf{P})\right]}_{\text{Markov chain parameters}} \underbrace{\left[P(\mathbf{Y}_T | \boldsymbol{\Phi}, \mathbf{S}_T) P(\boldsymbol{\Phi})\right]}_{\text{Structural parameters}} \quad (1.4)$$

Equation (1.3) follows from assumption (a), and equation (1.4) follows from assumption (b). Each block in equation (1.4) corresponds to a particular step of the MCMC algorithm. I describe each of the steps in detail in the following subsections.

Step 1: Sampling the regime indicators' parameters

Since I assume that the priors and likelihoods are independent across indicators, I can draw from their posterior distributions separately. In particular, let us consider the posterior of the parameters of the k^{th} indicator. Since each column of the matrix $\mathbf{G}^{(k)}$ is a vector of non-negative numbers that sum to unity, I will use the Dirichlet distribution, which turns out to be the natural conjugate. Let $\left(g_{1,s}^{(k)}, \ldots, g_{S,s}^{(k)}\right)$ be column s of $\mathbf{G}^{(k)}$. For $s = 1, \ldots, S$, my prior for column s is

$$\left(g_{1,s}^{(k)}, \ldots, g_{S,s}^{(k)}\right) \sim D\left(\alpha_{1,s}^{(k)}, \ldots, \alpha_{S,s}^{(k)}\right)$$

Assuming independent priors across columns, it follows that

$$P\left(\mathbf{G}^{(k)}\right) \propto \prod_{s=1}^{S} \prod_{s'=1}^{S} \left(g_{s',s}^{(k)}\right)^{\alpha_{s',s}^{(k)} - 1}$$

Let $\eta_{s',t}^{(k)} = 1$ if $Z_t^{(k)} = s'$, $\delta_{s,t} = 1$ if $S_t = s$, and let $H_{s',s}^{(k)} = \sum_{t=1}^{T} \eta_{s',t}^{(k)} \delta_{s,t}$, so $H_{s',s}^{(k)}$ counts the number of times that the regime is s and the indicator is s' in the sequence $\mathbf{S}_T$. From equation (1.4), the component of the conditional likelihood that depends on $\mathbf{G}^{(k)}$ is

$$P\left(\mathbf{Z}_T^{(k)}|\mathbf{G}^{(k)},\mathbf{S}_T\right) \propto \prod_{s=1}^{S}\prod_{s'=1}^{S}\left(g_{s',s}^{(k)}\right)^{H_{s',s}^{(k)}}$$

By Bayes' rule, the corresponding posterior distribution is

$$P\left(\mathbf{G}^{(k)}|\mathbf{Z}_T^{(k)},\mathbf{S}_T\right) \propto \prod_{s=1}^{S}\prod_{s'=1}^{S}\left(g_{s',s}^{(k)}\right)^{\alpha_{s',s}^{(k)}+H_{s',s}^{(k)}-1}$$

Thus, for each column s of $\mathbf{G}^{(k)}$, the posterior distribution is

$$\left(g_{1,s}^{(k)},\ldots,g_{S,s}^{(k)}|\mathbf{Z}_T^{(k)},\mathbf{S}_T\right) \sim D\left(\alpha_{1,s}^{(k)}+H_{1,s}^{(k)},\ldots,\alpha_{S,s}^{(k)}+H_{S,s}^{(k)}\right) \tag{1.5}$$

In the special case of uninformative indicator, $\mathbf{G}_{\cdot,s}^{(k)} = \mathbf{G}_{\cdot,s'}^{(k)}$ for all $s,s' = 1,\ldots S$, the above posterior distribution becomes

$$\left(g_1^{(k)},\ldots,g_S^{(k)}|\mathbf{Z}_T^{(k)},\mathbf{S}_T\right) \sim D\left(\alpha_1^{(k)}+\sum_{s=1}^{S}H_{1,s}^{(k)},\ldots,\alpha_S^{(k)}+\sum_{s=1}^{S}H_{S,s}^{(k)}\right) \tag{1.6}$$

Step 2: Sampling the initial and transition probabilities

Because the initial probabilities and each column of the transition matrix is a vector of non-negative numbers that sum to unity, I will use the Dirichlet distribution as in Step 1. Let $(p_{1,s},\ldots,p_{S,s})$ be column s of $\mathbf{P}$. For $s = 1,\ldots,S$, my priors for the initial and transition probabilities are

$$(\mu_1,\ldots,\mu_S) \sim D(\beta_1,\beta_2,\ldots,\beta_S)$$
$$(p_{1,s},\ldots,p_{S,s}) \sim D(\beta_{1,s},\beta_{2,s},\ldots,\beta_{S,s})$$

Assuming independent priors across columns of the transition probability matrix, it follows that

$$P(\boldsymbol{\mu},\mathbf{P}) \propto \left(\prod_{s=1}^{S}\mu_s^{\beta_s-1}\right)\left(\prod_{s=1}^{S}\prod_{s'=1}^{S}p_{s',s}^{\beta_{s',s}-1}\right)$$

From equation (1.4), the component of the conditional likelihood that depends on $\boldsymbol{\mu}$ and $\mathbf{P}$ is

$$P(\mathbf{S}_T|\boldsymbol{\mu},\mathbf{P}) \propto \left(\prod_{s=1}^{S}\mu_s^{\delta_{s,1}}\right)\left(\prod_{s=1}^{S}\prod_{s'=1}^{S}p_{s',s}^{T_{s',s}}\right)$$

where $T_{s',s} = \sum_{t=2}^{T}\delta_{s,t-1}\delta_{s',t}$, $\delta_{s,t} = 1$ if $S_t = s$ and 0 otherwise. In other words, $T_{s',s}$ is the total number of transition from regime s to regime s' in the sequence $\mathbf{S}_T$. By Bayes' rule, the corresponding posterior distribution is

$$P(\boldsymbol{\mu},\mathbf{P}|\mathbf{S}_T) \propto \left(\prod_{s=1}^{S}\mu_s^{\beta_s+\delta_{s,1}-1}\right)\left(\prod_{s=1}^{S}\prod_{s'=1}^{S}p_{s',s}^{\beta_{s',s}+T_{s',s}-1}\right)$$

Thus, the posterior distributions for the initial probabilities, $\boldsymbol{\mu}$, and for each column s of the transition matrix, $\mathbf{P}$, are

$$(\mu_1,\ldots,\mu_S|\mathbf{S}_T) \sim D(\beta_1+\delta_{1,1},\beta_2+\delta_{2,1},\ldots,\beta_S+\delta_{S,1}) \tag{1.7}$$

$$(p_{1,s},\ldots,p_{S,s}|\mathbf{S}_T) \sim D(\beta_{1,s}+T_{1,s},\beta_{2,s}+T_{2,s},\ldots,\beta_{S,s}+T_{S,s}) \tag{1.8}$$

Step 3: Sampling the structural parameters

To estimate structural parameters, I use independent priors across states. The priors are

$$P(\boldsymbol{\Phi}) = \prod_{s=1}^{S}P(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s) \tag{1.9}$$

From equation (1.4), the component of the conditional likelihood that depends on $\mathbf{\Phi}$ is

$$P(\mathbf{Y}_T|\mathbf{\Phi},\mathbf{S}_T) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{\frac{n}{2}}} \sum_{s=1}^{S} \delta_{s,t} \frac{|\mathbf{A}_{S_t}|}{|\mathbf{D}_{S_t}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{A}_{S_t}\mathbf{y}_t - \mathbf{B}_{S_t}\mathbf{x}_{t-1})' \mathbf{D}_{S_t}^{-1} (\mathbf{A}_{S_t}\mathbf{y}_t - \mathbf{B}_{S_t}\mathbf{x}_{t-1})\right\} \tag{1.10}$$

Let $\Delta(s) = \{t \in 1,\ldots,T : \delta_{st} = 1\}$ denote the set of dates for which the regime is s, and let $\mathbf{Y}_{\Delta(s)} = \{\mathbf{y}_t, \mathbf{x}_{t-1} : t \in \Delta(s)\}$ denote the observations of the endogenous variables in regime s. It follows that the conditional posterior distribution is

$$P(\mathbf{\Phi}|\mathbf{Y}_T,\mathbf{S}_T) \propto \prod_{s=1}^{S} P\left(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s|\mathbf{Y}_{\Delta(s)}\right) \tag{1.11}$$

where

$$\begin{aligned} P\left(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s|\mathbf{Y}_{\Delta(s)}\right) \propto & P(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s) \\ & \times \prod_{t\in\Delta(s)} \frac{|\mathbf{A}_s|}{|\mathbf{D}_s|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{A}_s\mathbf{y}_t - \mathbf{B}_s\mathbf{x}_{t-1})' \mathbf{D}_s^{-1} (\mathbf{A}_s\mathbf{y}_t - \mathbf{B}_s\mathbf{x}_{t-1})\right] \end{aligned} \tag{1.12}$$

Therefore, under the prior specifications and conditioning on the states and the transition probabilities, the estimation problem reduces to estimating S different sign-restricted SVAR models. Thus, we can decompose our observations into S separate groups according to their respective states and apply the algorithm of Baumeister & Hamilton (2015) to each group. The following provides a brief summary of their algorithm.

Fixing a specific state s, the objective is to simulate from $P\left(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s|\mathbf{Y}_{\Delta(s)}\right)$, the posterior distributions of $(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s)$. For estimation purpose, I decompose the joint prior on $(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s)$ as

$$P(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s) = P(\mathbf{A}_s)P(\mathbf{D}_s|\mathbf{A}_s)P(\mathbf{B}_s|\mathbf{A}_s,\mathbf{D}_s) \tag{1.13}$$

To facilitate computation of the posteriors, I allow arbitrary priors on parameters of $\mathbf{A}_s$ and use natural conjugates for the two conditional priors $P(\mathbf{D_s}|\mathbf{A_s})$ and $P(\mathbf{B}_s|\mathbf{A}_s,\mathbf{D}_s)$. Specifically, I use independent inverse-Gamma priors for the variances of the structural shocks

$$p\left(d_{ii,s}^{-1}|\mathbf{A}_s\right) \sim \Gamma(\kappa_{i,s},\tau_{i,s}) \tag{1.14}$$

$$P(\mathbf{D}_s|A_s) = \prod_{i=1}^{n} p\left(d_{ii,s}^{-1}|\mathbf{A}_s\right) \tag{1.15}$$

where $\Gamma(\kappa_{i,s},\tau_{i,s})$ denotes the Gamma distribution with shape parameter $\kappa_{i,s}$ and rate parameter $\tau_{i,s}$. For the lag parameters, I use multivariate normal priors and let them be independent across equations

$$P(\mathbf{b}_{i,s}|\mathbf{A}_s,\mathbf{D}_s) \sim N(\mathbf{m}_{i,s}, d_{ii,s}\mathbf{M}_{i,s}) \tag{1.16}$$

$$P(\mathbf{B}_s|\mathbf{A}_s,\mathbf{D}_s) = \prod_{i=1}^{n} P(\mathbf{b}_{i,s}|\mathbf{A}_s,\mathbf{D}_s) \tag{1.17}$$

where $N(\mathbf{m}_{i,s}, d_{ii,s}\mathbf{M}_{i,s})$ denotes the multivariate normal distribution with mean $\mathbf{m}_{i,s}$ and covariance matrix $d_{ii,s}\mathbf{M}_{i,s}$. The joint prior is

$$P(\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s) = P(\mathbf{A}_s)\prod_{i=1}^{n} p\left(d_{ii,s}^{-1}|\mathbf{A}_s\right)\prod_{i=1}^{n} P(\mathbf{b}_{i,s}|\mathbf{A}_s,\mathbf{D}_s) \tag{1.18}$$

The component of the conditional likelihood that depends on $\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s$ is

$$P\left(\mathbf{Y}_{\Delta(s)}|\mathbf{A}_s,\mathbf{B}_s,\mathbf{D}_s\right) \propto \prod_{t\in\Delta(s)} \frac{|\mathbf{A}_s|}{|\mathbf{D}_s|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{A}_s\mathbf{y}_t - \mathbf{B}_s\mathbf{x}_{t-1})'\mathbf{D}_s^{-1}(\mathbf{A}_s\mathbf{y}_t - \mathbf{B}_s\mathbf{x}_{t-1})\right\} \tag{1.19}$$

Given the priors and the likelihood, the posteriors are computed by Bayes' rule and summarized in the following proposition.

Proposition. *Let the priors be given as in* (1.13)-(1.18) *and the likelihood function be given as in* (1.19). *Moreover, let* $\mathbf{a}_{i,s}'$ *denote the i-th row of* $\mathbf{A}_s$, $\mathbf{L}_{i,s}$ *denote the Cholesky factor of*

$\mathbf{M}_{i,s}^{-1} = \mathbf{L}_{i,s}\mathbf{L}_{i,s}'$, *and* $T_{\Delta(s)}$ *be the total number of elements in the set* $\Delta(s)$*. Then, the posteriors are*

$$P\left(\mathbf{A}_s, \mathbf{B}_s, \mathbf{D}_s | \mathbf{Y}_{\Delta(s)}\right) = P\left(\mathbf{A}_s | \mathbf{Y}_{\Delta(s)}\right) \prod_{i=1}^{n} p\left(d_{ii,s}^{-1} | \mathbf{A}_s, \mathbf{Y}_{\Delta(s)}\right) \prod_{i=1}^{n} P\left(\mathbf{b}_{i,s} | \mathbf{A}_s, \mathbf{D}_s, \mathbf{Y}_{\Delta(s)}\right) \tag{1.20}$$

with

$$P\left(\mathbf{A}_s | \mathbf{Y}_{\Delta(s)}\right) \propto P\left(\mathbf{A}_s\right) \left|\det\left(\mathbf{A}_s\right)\right|^{T_{\Delta(s)}} \prod_{i=1}^{n} \left\{ \frac{\left|\mathbf{M}_{i,s}^*\right|^{\frac{1}{2}}}{\left|\mathbf{M}_{i,s}\right|^{\frac{1}{2}}} \frac{\tau_{i,s}^{\kappa_{i,s}}}{\Gamma\left(\kappa_{i,s}\right)} \frac{\Gamma\left(\kappa_{i,s}^*\right)}{\left(\tau_{i,s}^*\right)^{\kappa_{i,s}^*}} \right\} \tag{1.21}$$

$$p\left(d_{ii,s}^{-1} | \mathbf{A}_s, \mathbf{Y}_{\Delta(s)}\right) \sim \Gamma\left(\kappa_{i,s}^*, \tau_{i,s}^*\right) \tag{1.22}$$

$$P\left(\mathbf{b}_{i,s} | \mathbf{A}_s, \mathbf{D}_s, \mathbf{Y}_{\Delta(s)}\right) \sim N\left(\mathbf{m}_{i,s}^*, d_{ii,s}\mathbf{M}_{i,s}^*\right) \tag{1.23}$$

$$\kappa_{i,s}^* = \kappa_{i,s} + \frac{T_{\Delta(s)}}{2} \tag{1.24}$$

$$\tau_{i,s}^* = \tau_{i,s} + \frac{\zeta_{i,s}^*}{2} \tag{1.25}$$

$$\zeta_{i,s}^* = \left(\tilde{\mathbf{Y}}_{i,s}'\tilde{\mathbf{Y}}_{i,s}\right) - \left(\tilde{\mathbf{Y}}_{i,s}'\tilde{\mathbf{X}}_{i,s}\right)\left(\tilde{\mathbf{X}}_{i,s}'\tilde{\mathbf{X}}_{i,s}\right)^{-1}\left(\tilde{\mathbf{X}}_{i,s}'\tilde{\mathbf{Y}}_{i,s}\right) \tag{1.26}$$

$$\mathbf{m}_{i,s}^* = \left(\tilde{\mathbf{X}}_{i,s}'\tilde{\mathbf{X}}_{i,s}\right)^{-1}\left(\tilde{\mathbf{X}}_{i,s}'\tilde{\mathbf{Y}}_{i,s}\right) \tag{1.27}$$

$$\mathbf{M}_{i,s}^* = \left(\tilde{\mathbf{X}}_{i,s}'\tilde{\mathbf{X}}_{i,s}\right)^{-1} \tag{1.28}$$

$$\tilde{\mathbf{Y}}_{i,s} = \begin{bmatrix} \mathbf{y}_1'\mathbf{a}_i\delta_{s,1} & \ldots & \mathbf{y}_T'\mathbf{a}_i\delta_{s,T} & \mathbf{m}_{i,s}'\mathbf{L}_{i,s} \end{bmatrix}' \tag{1.29}$$

$$\tilde{\mathbf{X}}_{i,s} = \begin{bmatrix} \mathbf{x}_0'\delta_{s,1} & \ldots & \mathbf{x}_{T-1}'\delta_{s,T} & \mathbf{L}_{i,s} \end{bmatrix}' \tag{1.30}$$

A random-walk Metropolis-Hasting algorithm is used to draw from $P\left(\mathbf{A}_s | \mathbf{Y}_{\Delta(s)}\right)$. Let $\boldsymbol{\alpha}_s$ collect all unknown elements of $\mathbf{A}_s$, the target function is

$$\begin{aligned} q\left(\boldsymbol{\alpha}_s\right) = \ln\left(\boldsymbol{\alpha}_s\right) + T_{\Delta(s)} \ln\left|\det\left(\mathbf{A}_s\right)\right| + \sum_{i=1}^{n} \left\{ \frac{1}{2} \ln\left|\mathbf{M}_{i,s}^*\right| + \kappa_{i,s} \ln\left(\tau_{i,s}\right) + \ln\left[\Gamma\left(\kappa_{i,s}^*\right)\right] \right\} \\ - \sum_{i=1}^{n} \left\{ \frac{1}{2} \ln\left|\mathbf{M}_{i,s}\right| + \ln\left[\Gamma\left(\kappa_{i,s}\right)\right] + \kappa_{i,s}^* \ln\left(\tau_{i,s}^*\right) \right\} \end{aligned} \tag{1.31}$$

where $\kappa_{i,s}^*$, $\tau_{i,s}^*$, and $\mathbf{M}_{i,s}^*$, are given in (1.24), (1.25), and (1.28) respectively. The algorithm will be more efficient with a good approximation to the posterior distribution. Since the hidden states are unknown *ex ante*, the initial guess, $\hat{\boldsymbol{\alpha}}$, is formed by maximizing equation (1.31) over the entire sample. Then, the negative of the Hessian matrix, $\hat{\boldsymbol{\Lambda}}$, is calculated, and its Cholesky factor, $\hat{\boldsymbol{P}}_{\boldsymbol{\Lambda}}$, is used to approximate the scale of the posterior distribution. Next, let $\boldsymbol{\alpha}_s^{old}$ be the draw of $\boldsymbol{\alpha}_s$ from the previous iteration, to draw $\boldsymbol{\alpha}_s^{new}$, first generate a candidate draw from $\tilde{\boldsymbol{\alpha}}_s = \boldsymbol{\alpha}_s^{old} + \omega\left(\hat{\boldsymbol{P}}_{\boldsymbol{\Lambda}}^{-1}\right)' \boldsymbol{v}$, where ω is a tuning parameter and $\boldsymbol{v}$ is a symmetric proposal distribution, such as the Student t distribution. If $q\left(\tilde{\boldsymbol{\alpha}}_s\right) > q\left(\boldsymbol{\alpha}_s^{old}\right)$, set $\boldsymbol{\alpha}_s^{new} = \tilde{\boldsymbol{\alpha}}_s$. Otherwise, set $\boldsymbol{\alpha}_s^{new} = \boldsymbol{\alpha}_s^{old}$ with probability $1 - \exp\left[q\left(\tilde{\boldsymbol{\alpha}}_s\right) - q\left(\boldsymbol{\alpha}_s^{old}\right)\right]$.

This step differs from that of Baumeister & Hamilton (2015) in two main ways. First, because the Metropolis-Hasting algorithm is embedded within the Gibbs sampler, only *a single draw* of $\mathbf{A}_s$ is needed in each Gibbs iteration.[9] Second, the target for $P\left(\mathbf{A}_s|\mathbf{Y}_{\Delta(s)}\right)$, equation (1.21), does not involve an estimate of the reduced-form covariance matrix of each state.[10] This target is more operational because the algorithm might be trapped in one specific state in early iterations, and hence the reduced-form covariance matrices of other states cannot be calculated, forcing the algorithm to stop prematurely.

Lastly, conditioning on the new draw for $\mathbf{A}_s$, $\mathbf{D}_s$ and $\mathbf{B}_s$ are sampled from their respective natural conjugates. Specifically, equation (1.22) is used to draw $d_{ii,s}^{-1}$ from $p\left(d_{ii,s}^{-1}|\mathbf{A}_s, \mathbf{Y}_{\Delta(s)}\right)$ for $i = 1, \ldots, n$, and then, equation (1.23) is used to draw $\mathbf{b}_{i,s}$, the ith row of $\mathbf{B}_s$, from $P\left(\mathbf{b}_{i,s}|\mathbf{A}_s, \mathbf{D}_s, \mathbf{Y}_{\Delta(s)}\right)$ for $i = 1, \ldots, n$.

Step 4: Sampling the hidden states

Conditioning on the transition probabilities and the structural parameters, the sequence $\mathbf{S}_T$ is sampled by the multi-move Gibbs sampler as described in Chib (1996), Kim & Nelson (1999). Let $\boldsymbol{\theta}$ denote all unknown parameters, that is $\boldsymbol{\theta} = \left(\mathrm{vec}\left(\mathbf{G}\right)', \boldsymbol{\mu}', \mathrm{vec}\left(\mathbf{P}\right)', \mathrm{vec}\left(\boldsymbol{\Phi}\right)'\right)'$, the

[9]Carlin & Louis (2010) notes that convergence of the Gibbs sampler is achieved for any number of Metropolis-Hasting sub-iterations, hence one sub-iteration is often adopted in practice.

[10]This target comes from equation (60) in Baumeister & Hamilton (2014).

sampler uses the following equations

$$P\left(\mathbf{S}_T|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)=P\left(S_T|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)\prod_{t=1}^{T-1}P\left(S_t|S_{t+1},\boldsymbol{\theta}^{(m+1)},\mathbf{W}_t\right) \tag{1.32}$$

$$P\left(S_t|S_{t+1},\boldsymbol{\theta}^{(m+1)},\mathbf{W}_t\right)\propto P\left(S_{t+1}|S_t\right)P\left(S_t|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_t\right) \tag{1.33}$$

Equation (1.32) decomposes the joint conditional posterior distribution of $\mathbf{S}_T$ into a form that is applicable to the forward-backward algorithm. The strategy is to first run the Hamilton filter, as described in Appendix A.1, to estimate the filtered probabilities of the hidden states, $P\left(S_t|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_t\right)$. Then, we iterate backward from the last value of this sequence, which is $P\left(S_T|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)$.

Specifically, to generate one realization for $S_T^{(m+1)}$, we first draw a random number, u, from the uniform distribution, $U(0,1)$. If $u<P\left(S_T=1|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)$, we will set $S_T^{(m+1)}=1$. If $P\left(S_T=1|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)<u<P\left(S_T=1|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)+P\left(S_T=2|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)$, we will set $S_T^{(m+1)}=2$, and so on. After having the value of $S_T^{(m+1)}$, we use equation (1.33) to simulate $S_{T-1}^{(m+1)}$. First, we calculate

$$P\left(S_{T-1}|S_T=S_T^{(m+1)},\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T\right)=\frac{P\left(S_T=S_T^{(m+1)}|S_{T-1}\right)P\left(S_{T-1}|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_{T-1}\right)}{\sum_{j=1}^{N}P\left(S_T|S_{T-1}=j\right)P\left(S_{T-1}=j|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_{T-1}\right)} \tag{1.34}$$

Then, as before, we use a random number generated from a uniform distribution to generate $S_{T-1}^{(m+1)}$. Continuing this process, we will get a realization for the hidden states, $\mathbf{S}_T^{(m+1)}$, which come from their conditional posterior distribution $\mathbf{S}_T|\boldsymbol{\theta}^{(m+1)},\mathbf{W}_T$.

Step 5: Permutation sampler

Finally, to deal with the label-switching problem, I use the permutation sampler as described in Frühwirth-Schnatter (2006). To implement the sampler, after having a new draw, we simply switch their current labeling based on a random permutation among all $S!$ possible permutations of the labeling. Denote a realization of the random permutation as $\{\rho(1), \rho(2), \ldots, \rho(S)\}$, we will

1. Apply the permutation to the regime indicator parameters, $\left(\mathbf{G}^{(k)}\right)^{(m+1)}$, by substituting $\mathbf{G}^{(k)}_{\cdot,s}$ with $\mathbf{G}^{(k)}_{\cdot,\rho(s)}$ for $s = 1,2,\ldots,S$, and $k = 1,2,\ldots,K$.

2. Apply the permutation to the initial probabilities, $\boldsymbol{\mu}^{(m+1)}$, by substituting $\mu_s^{(m+1)}$ with $\mu_{\rho(s)}^{(m+1)}$ for $s = 1,2,\ldots,S$.

   Apply the permutation to the transition matrix, $\mathbf{P}^{(m+1)}$, by substituting $p_{s,s'}^{(m+1)}$ with $p_{\rho(s),\rho(s')}^{(m+1)}$ for $s,s' = 1,2,\ldots,S$.

3. Apply the permutation to the structural parameters, $\boldsymbol{\Phi}^{(m+1)}$, by substituting $\Phi_s^{(m+1)}$ with $\Phi_{\rho(s)}^{(m+1)}$ for $s = 1,2,\ldots,S$.

4. Apply the permutation to the hidden states, $\mathbf{S}_T^{(m+1)}$ by substituting $S_t^{(m+1)}$ with $\rho\left(S_t^{(m+1)}\right)$ for $t = 1,2,\ldots,T$.

1.2.4 Structural Impulse-Response Functions

Given parameters $(\mathbf{A}_s, \mathbf{B}_s, \mathbf{D}_s)$, we can calculate the SIRFs associated with regime s. To that end, define

$$\mathbf{F}_s = \begin{bmatrix} \mathbf{A}_s^{-1}\mathbf{B}_{1,s} & \mathbf{A}_s^{-1}\mathbf{B}_{2,s} & \ldots & \mathbf{A}_s^{-1}\mathbf{B}_{p-1,s} & \mathbf{A}_s^{-1}\mathbf{B}_{p,s} \\ \mathbf{I}_n & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

Let $\mathcal{F}_t$ denote the information set at time t, the dynamic causal effects of structural shocks in regime s at horizon h is given by

$$\mathbf{H}_{h,s} \equiv \frac{\partial \mathbb{E}\left(\mathbf{y}_{t+h} \middle| S_t = s, S_{t+1} = s, \ldots, S_{t+h} = s, \mathcal{F}_t\right)}{\partial \mathbf{u}_t'} = \boldsymbol{\Psi}_{h,s} \mathbf{A}_s^{-1} \tag{1.35}$$

where $\boldsymbol{\Psi}_{h,s}$ is found from the first n rows and columns of $\mathbf{F}_s^h$.

Some papers, such as Ramey & Zubairy (2018) and Lyu & Noh (2018), look at the SIRFs conditional only on starting in regime s

$$\mathbf{H}_{h,s}^{*} \equiv \frac{\partial \mathbb{E}\left(\mathbf{y}_{t+h} \middle| S_t = s\right)}{\partial \mathbf{u}_t'}$$

This object is the weighted average of (1.35) across different sequences of states. In the special case of two hidden states, $\mathbf{H}_{h,s}$, evaluated at each state s, are the lower and upper bounds for $\mathbf{H}_{h,s}^{*}$. Therefore, the similarity of $\mathbf{H}_{h,s}$ across regimes will be persuasive evidence against asymmetry.

1.2.5 Bayesian model comparison

Using Bayesian model comparison, we can answer two important questions: (1) Is there a regime change in the data? and (2) Is the qualitative information for the hidden states useful? The first question is challenging for frequentists since models with different regimes are not necessarily nested.[11] However, Bayesians can still solve the problem by comparing models with different regimes. Indeed, Koop & Potter (1999) points out several advantages of Bayesian methods over their frequentist counterparts in nonstandard inference. Similarly, the second question can be answered by a Bayesian model comparison between a model with the qualitative information and another one where the qualitative information is restricted to be uninformative.

Generally, a Bayesian model comparison between model 1 and 2 is done by calculating the posterior odds ratio, showing which model is more likely now that we have seen the data.

[11] See Hamilton (2016) for a review of frequentist methods in determining the number of regimes.

This can be written as the product of Bayes factor and the prior odds ratio.

$$\frac{P(M_2|\mathbf{W}_T)}{P(M_1|\mathbf{W}_T)} = \frac{P(\mathbf{W}_T|M_2)}{P(\mathbf{W}_T|M_1)}\frac{P(M_2)}{P(M_1)} \tag{1.36}$$

Suppose before seeing the data, we assume model 1 is as likely as model 2, then our prior odds ratio (i.e. $P(M_2)/P(M_1)$) will be one and equation (1.36) simplifies to

$$\frac{P(M_2|\mathbf{W}_T)}{P(M_1|\mathbf{W}_T)} = \frac{P(\mathbf{W}_T|M_2)}{P(\mathbf{W}_T|M_1)} \tag{1.37}$$

The quantity on the right-hand side of equation (1.37) is the so-called Bayes factor, which is the ratio of two marginal likelihoods. The Bayes factor between model 2 and model 1 is denoted as

$$B_{21} = \frac{P(\mathbf{W}_T|M_2)}{P(\mathbf{W}_T|M_1)}$$

The Bayes factor is the ratio of two marginal likelihoods, which are high dimensional integrals, and hence its computation is technically challenging. Let N_0 be the size of the posterior draws after discarding the burn-in sample. We estimate the marginal likelihood for each model from the posterior draws as

$$\hat{P}(\mathbf{W}_T|M_l) = \left[\frac{1}{N_0}\sum_{n_0=1}^{N_0}\frac{f(\boldsymbol{\theta}^{n_0})}{P(\mathbf{W}_T|\boldsymbol{\theta}^{n_0})P(\boldsymbol{\theta}^{n_0})}\right]^{-1} \quad \text{for } l = 1,2 \tag{1.38}$$

where $f(.)$ is chosen to closely approximates the posterior distribution while leaving out extreme values from the simulation. Geweke (1999) proposes to use the truncated normal distribution while Sims, Waggoner & Zha (2008) suggests a more sophisticated choice to deal with non-Gaussian posterior. Details for both methods are given in Appendix A.2.

1.3 Empirical application: Asymmetric effects of monetary policy

Is conventional monetary policy more effective during recessions or expansions? To illustrate my method, I revisit the literature on asymmetric effects of conventional monetary policy over the business cycle by estimating a trivariate 2-state Markov-switching SVAR model identified by sign restrictions. I allow all model parameters to depend on a hidden state governed by an exogenous 2-state Markov chain while using the sign restrictions developed by Baumeister & Hamilton (2018) to identify the regime-dependent dynamic causal effects. I first use symmetric priors in both regimes, then I further check for robustness by augmenting the baseline model with the NBER recession indicator and asymmetric priors about structural parameters. In the following subsections, I describe the data, models, and priors in detail.

1.3.1 Data Description

The data in this study are publicly available and are downloaded from FRED. I provide a detailed description of the data and their sources in Appendix A.3. I use five *quarterly* time series: real GDP, real potential GDP, Personal Consumption Expenditures deflator (PCE deflator), effective federal funds rate, and NBER based recession indicator. I calculate the output gap as the log difference between real and potential GDP, and inflation rate as the Y/Y change of the PCE deflator. The sample period is 1954Q3 to 2007Q4. Figure 1.1 plots the transformed time series of macroeconomic variables. This figure shows that all macroeconomic variables and are less volatile after the 1980s, which is consistent with Stock & Watson (2002)'s documentation of the Great Moderation. Table 1.1 contains the summary statistics of my variables.

Table 1.1: **Summary statistics**. This table shows the summary statistics of variables in the empirical application. Descriptions of the data and their availability are explained in the text. There are 214 quarterly observations for each variable in the period between 1954Q3 and 2007Q4. The units are all in percentage points.

Summary statistics				
Variables	**Mean**	**Standard Deviations**	**Minimum**	**Maximum**
Output gap	-0.5	2.2	-7.6	5.5
Inflation rates	3.5	2.3	0.6	10.9
Fed Fund rates	5.8	3.3	0.9	17.8

1.3.2 A regime-switching sign-restricted SVAR model of monetary policy

Here I introduce the trivariate model of output gap, inflation, and nominal Fed funds rate in Baumeister & Hamilton (2018). My main innovation is to allow the structural parameters to be different across regimes. Let y_t be the output gap, π_t be the inflation rate, r_t be the nominal Fed fund rate, the model consists of the following structural equations

1. The Phillips curve:
$$y_t = k^s_{S_t} + \alpha^s_{S_t}\pi_t + \left[\mathbf{b}^s_{S_t}\right]'\mathbf{x}_{t-1} + u^s_t \tag{1.39}$$

2. The aggregate demand equation:
$$y_t = k^d_{S_t} + \beta^d_{S_t}\pi_t + \gamma^d_{S_t} r_t + \left[\mathbf{b}^d_{S_t}\right]'\mathbf{x}_{t-1} + u^d_t \tag{1.40}$$

3. The monetary policy rule:
$$r_t = k^m_{S_t} + (1-\rho_{S_t})\psi^y_{S_t} y_t + (1-\rho_{S_t})\psi^\pi_{S_t}\pi_t + \rho_{S_t} r_{t-1} + \left[\mathbf{b}^m_{S_t}\right]'\mathbf{x}_{t-1} + u^m_t \tag{1.41}$$

4. The Markov transition probabilities:

$$\mathbf{P} = \begin{bmatrix} P(S_t = 1|S_{t-1} = 1) & P(S_t = 1|S_{t-1} = 2) \\ P(S_t = 2|S_{t-1} = 1) & P(S_t = 2|S_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{1.42}$$

Together, equations (1.39), (1.40), (1.41), and (1.42) constitute the baseline model.[12] I impose four sign restrictions on the structural parameters and two sign restrictions on the elements of $\mathbf{A}_s^{-1}$. In particular, I assume the following sign restrictions hold for any state $s \in \{1,2\}$: (1) the Phillips curve is downward sloping $(\alpha_s^s > 0)$, (2) raising interest rate will not stimulate aggregate demand $\left(\gamma_s^d < 0\right)$, (3) the Fed will raise the interest rate when inflation is higher than its target$\left(\psi_s^y > 0\right)$ or output gap is higher than its potential $(\psi_s^\pi > 0)$, and (4) the Fed wants to increase its interest rate smoothly over time $(0 < \rho_s < 1)$.

To incorporate priors on the elements of $\mathbf{A}_s^{-1}$, I first define $h_{1,s} = \beta_s^d + \gamma_s^d (1 - \rho_s) \psi_s^\pi$ and $h_{2,s} = \dfrac{\alpha_s^s \gamma_s^d}{\alpha_s^s - \beta_s^d}$. Then, I use the asymmetric t distributions to nudge those quantities to the region of the parameter spaces where both of them are negative. Therefore, my expected signs for the elements of $\mathbf{A}_s^{-1}$ are

$$\mathbf{A}_s^{-1} = \begin{bmatrix} + & + & - \\ - & + & - \\ ? & + & + \end{bmatrix}$$

Priors for the transition probabilities

I use the Dirichlet(1,1) distribution for each column of the transition probabilities matrix **P**. It is equivalent to a uniform distribution over $(0,1)$ as prior for each element of that matrix. As in the general formulation, I assume independent priors between columns. Intuitively, my priors are that every number between 0 and 1 is equally plausible for the transition probabilities and that the probability of moving to state 1 does not depend on the probability of moving to state 2.

[12] For simplicity, I set the initial probabilities to be 0.5 for each state.

Formally, the priors are

$$P(\mathbf{P}) = P(p_{11}, p_{21})\, P(p_{12}, p_{22})$$
$$(p_{11}, p_{21}) \sim D(1,1)$$
$$(p_{12}, p_{22}) \sim D(1,1)$$

For the structural parameters, I use independent and symmetric priors. In the following, I will describe my priors for structural parameters in a state $s \in \{1,2\}$.

Priors for contemporaneous coefficients

Structural parameters in the general formulation are reflected in $\mathbf{A}_s$, $\mathbf{B}_s$, and $\mathbf{D}_s$. My priors for elements of $\mathbf{A}_s$ are

$$P(\mathbf{A}_s) = c_s p(\alpha_s^s)\, p\left(\beta_s^d\right) p\left(\gamma_s^d\right) p(\psi_s^y)\, p(\psi_s^\pi)\, p(\rho_s)\, (p(h_{1,s}))^{\zeta_{h_{1,s}}} (p(h_{2,s}))^{\zeta_{h_{2,s}}}$$

where c_s is the integrating constant that needs to be accounted for when calculating the marginal likelihood.[13] The two parameters $\xi_{h_{1,s}}$ and $\xi_{h_{2,s}}$ reflect how strong the priors about elements of $\mathbf{A}_s^{-1}$ are relative to those of other parameters in $\mathbf{A}_s$. In my application, I set both to be 1. I follow Baumeister & Hamilton (2018) in setting prior values for elements of $\mathbf{A}_s$, thus I do not repeat their economic motivations in this paper.[14] The following will describe my priors for $\mathbf{B}_s, \mathbf{D}_s$ conditional on $\mathbf{A}_s$.

[13]I estimate c_s by $\hat{c}_s = \left[\frac{1}{500,000} \sum_{m=1}^{500,000} p\left(h_{1,s}^{(m)}\right)^{\zeta_{h_{1,s}}} p\left(h_{1,s}^{(m)}\right)^{\zeta_{h_{1,s}}}\right]^{-1}$ where $\left(h_{1,s}^{(m)}\right)$ and $\left(h_{2,s}^{(m)}\right)$ are calculated with draws from the prior distributions.

[14]Belongia & Ireland (2019) presents another approach for setting priors of contemporaneous structural parameters in SVAR models of monetary policy.

Priors for covariance matrix

The priors for elements of the covariance matrix are set to be independent from each other.

$$P(\mathbf{D}_s|\mathbf{A}_s) = \prod_{i=1}^{3} p\left(d_{ii,s}^{-1}|\mathbf{A}_s\right)$$

where each of the element $d_{ii,s}$ follows an inverse-Gamma distribution

$$p\left(d_{ii,s}^{-1}|\mathbf{A}_s\right) = \begin{cases} \dfrac{\tau_{i,s}^{\kappa_i}}{\Gamma(\kappa_{i,s})} \quad \left(d_{ii,s}^{-1}\right)^{k_{i,s}-1} \exp\left(-\tau_{i,s} d_{ii,s}^{-1}\right) & \text{for } d_{ii,s}^{-1} \geq 0, \\ 0 \qquad \text{otherwise,} & \end{cases}$$

In this specification, $\dfrac{\kappa_{i,s}}{\tau_{i,s}}$ is the prior mean for $d_{ii,s}^{-1}$, and $\dfrac{\kappa_{i,s}}{\tau_{i,s}^2}$ is the prior variance.

Since the priors should reflect the scale of the data, I first fit three univariate $AR(4)$ model

$y_t = \beta_{10} + \sum_{i=1}^{4} \beta_{1i} y_{t-i} + e_{1t}$

$\pi_t = \beta_{20} + \sum_{i=1}^{4} \beta_{2i} \pi_{t-i} + e_{2t}$

$r_t = \beta_{30} + \sum_{i=1}^{4} \beta_{3i} r_{t-i} + e_{3t}$

Then, I use the fitted residuals from those regressions, $\hat{e}_t = \begin{bmatrix} \hat{e}_{1t} & \hat{e}_{2t} & \hat{e}_{3t} \end{bmatrix}'$, to estimate the sample covariance matrix

$\hat{\mathbf{S}} = \dfrac{1}{T} \sum_{t=1}^{T} \hat{e}_t \hat{e}_t'$

Note that the sample covariance matrix is estimated using the entire sample regardless of the state, and hence it is fixed during every iteration of the Gibbs sampler.

Finally, I set $\kappa_{i,s} = 2$ for all i, which gives my prior a weight that equals to four observation of the data in the posterior. Next, I set the prior mean for $d_{ii,s}^{-1}$ to be the reciprocal of the i^{th} diagonal element of $\mathbf{A}_s \hat{\mathbf{S}} \mathbf{A}_s'$. In sum, the priors for elements of the covariance matrix will be

$$p\left(d_{ii,s}^{-1}|\mathbf{A}_s\right) \sim \Gamma\left(2, 2\mathbf{a}_{i,s}' \hat{\mathbf{S}} \mathbf{a}_{i,s}\right)$$

Priors for lag coefficients

The SVAR model has four lags whose priors are set so that their counterparts for the reduced-form coefficients are consistent with the Minnesota priors. Specifically, the priors for lag coefficients are assumed to be independent across equations. That is

$$P(\mathbf{B}_s|\mathbf{A}_s,\mathbf{D}_s)=\prod_{i=1}^{3}P(\mathbf{b}_{i,s}|\mathbf{A}_s,\mathbf{D}_s)$$

where $P(\mathbf{b}_{i,s}|\mathbf{A}_s,\mathbf{D}_s)\sim N(\mathbf{m}_{i,s},d_{ii,s}\mathbf{M}_i)$. For the mean, I set $\mathbf{m}_{i,s}(\alpha)=\eta'\mathbf{a}_{i,s}$ where $\eta=\begin{bmatrix}\mathbf{I}_3 & \mathbf{0}_{3\times 10}\end{bmatrix}$.[15] And for the covariance matrix $\mathbf{M}_i$, let $\sqrt{s_{ii}}$ be the estimated standard deviation of the $AR(4)$ that fits to variable i, we define

$$\mathbf{v}_1'=\begin{bmatrix}\frac{1}{1^{2\lambda_1}} & \frac{1}{2^{2\lambda_1}} & \frac{1}{3^{2\lambda_1}} & \frac{1}{4^{2\lambda_1}}\end{bmatrix}$$
$$\mathbf{v}_2'=\begin{bmatrix}s_{11}^{-1} & s_{22}^{-1} & s_{33}^{-1}\end{bmatrix}$$
$$\mathbf{v_3}=\lambda_0^2\begin{bmatrix}\mathbf{v}_1\mathbf{v}_2\\ \lambda_3^2\end{bmatrix}$$

Then, $\mathbf{M}_i$ will be the diagonal matrix whose row r column r element is the r^{th} element of $\mathbf{v}_3$: $M_{i,rr}=v_{3r}$. Intuitively, we are letting coefficients on higher lags to shrink to zero by setting decreasing values for diagonal elements of $\mathbf{M}_i$. The hyper-parameter λ_0 captures our confidence in the priors: a higher value implies a higher variance and less confidence. The hyper-parameter λ_1 governs how quickly the coefficients shrink to zero. And the hyper-parameter λ_3 describes our confidence in the prior for the constant: the higher the value, the less confidence we have. I set $\lambda_0=0.2$, $\lambda_1=1$, and $\lambda_3=100$.

Also, as seen in the monetary policy rule, I restrict the third element of that equation to be

[15]Note that if the reduced-form lag coefficients (Φ) follow the Minnesota prior (meaning $\mathbb{E}(\mathbf{\Phi})=\eta$). Then, $\mathbb{E}(\mathbf{B}|\mathbf{A})=\mathbb{E}(\mathbf{A}\Phi|\mathbf{A})=\mathbf{A}\mathbb{E}(\Phi|\mathbf{A})=\eta\mathbf{A}$. Thus, if the priors for the reduced-form lag coefficients are the Minnesota prior, the priors for the lag coefficients in the structural model will be normal with mean $\mathbf{m}_i(\alpha)=\eta'\mathbf{a}_i$.

close to ρ_s. The restriction can be described as a prior on the third element of $\mathbf{b}_{3,s}$ as follow

$$\rho_s = \mathbf{I}_{13}^{(3)} \mathbf{b}_{3,s} + v_3 \tag{1.43}$$

where $v_3 \sim N(0, d_{33,s} V_3)$ and $\mathbf{I}_{13}^{(3)}$ is the third row of $\mathbf{I}_{13}$. The variance parameter, V_3, reflects the strength of our priors because the smaller V_3 is, the more likely that parameter is close to ρ_s. I set $V_3 = 0.1$. To estimate the model with the restriction (1.43), we apply the same algorithm as described in Baumeister & Hamilton (2015) with some modifications for $\tilde{\mathbf{Y}}_{i,s}$ and $\tilde{\mathbf{X}}_{i,s}$. In this case, the new $\tilde{\mathbf{Y}}_{3,s}$ and $\tilde{\mathbf{X}}_{3,s}$ are

$$\tilde{\mathbf{Y}}_{3,s} = \begin{bmatrix} \mathbf{y}_1' \mathbf{a}_{3,s} \delta_{s,1} & \ldots & \mathbf{y}_T' \mathbf{a}_{3,s} \delta_{s,T} & \mathbf{m}_{3,s}' \mathbf{L}_{3,s} & \frac{\rho_s}{\sqrt{V_3}} \end{bmatrix}$$
$$\tilde{\mathbf{X}}_{3,s} = \begin{bmatrix} \mathbf{x}_0' \delta_{s,1} & \ldots & \mathbf{x}_{T-1}' \delta_{s,T} & \mathbf{L}_{3,s} & \frac{\mathbf{I}_{13}^{(3)}}{\sqrt{V_3}} \end{bmatrix}$$

Table 1.2 summarizes the priors for each parameter together with its corresponding hyper-parameters. Combining the above prior distributions with the Gaussian likelihood function, I simulate the posterior distribution of this sign-restricted 2-state Markov-switching SVAR model by my MCMC algorithm. I tune the Metropolis-Hasting acceptance rate close to 23.4% as suggested in the MCMC literature (Roberts & Rosenthal (2001)).[16] In all of the applications, I simulate 1,000,000 draws and discard the first 500,000 draws. For convergence diagnostics, I visually inspect the trace plots and autocorrelation functions of all parameters. As an additional robustness check, I also apply the separated partial means test, as described in Geweke (1999) and Geweke (2005), to the MCMC outputs. The overall results suggest that the algorithm does have good mixing properties.[17]

[16]To tune the Metropolis-Hasting algorithm that generates $\mathbf{A}_s$, I use the procedure in subsection 5.4.2 in Baumeister & Hamilton (2015). Also, I randomly generate initial values for other parameters from their priors and initial values for the hidden states from the Bernoulli distribution.

[17]I use the first 10% and the last 50% of the MCMC outputs for the separated partial means tests. Out of 100 parameters of the baseline model, the test rejects the null hypothesis of equal means for only one parameter in the lag coefficient matrix. Diagnostic results for other MCMC outputs are qualitatively similar.

Table 1.2: **Prior distributions of the baseline model: Symmetric priors**. This table shows the prior distribution of $\mathbf{A}_s$, $\mathbf{D}_s$, $\mathbf{B}_s$, and $\mathbf{P}$ together with their hyper-parameters. For Student t and Asymmetric t distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

Parameter	Meaning	Location	Scale	Skew	Sign restriction
Priors for contemporaneous coefficients of $\mathbf{A}_s$ and elements of $\mathbf{A}_s^{-1}$					
Student t distribution with 3 degrees of freedom					
α_s^s	Effect of π on supply	2	0.4	–	$\alpha_s^s \geq 0$
β_s^d	Effect of π on demand	0.75	0.4	–	None
γ_s^d	Effect of r on demand	-1	0.4	–	$\gamma_s^d \leq 0$
ψ_s^y	Fed response to y	0.5	0.4	–	$\psi_s^y \geq 0$
ψ_s^π	Fed response to π	1.5	0.4	–	$\psi_s^\pi \geq 0$
Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$					
ρ_s	Interest rate smoothing	0.5	0.2	–	$0 \leq \rho_s \leq 1$
Asymmetric t distribution with 3 degrees of freedom					
$h_{1,s}$	Part of $\det(\mathbf{A}_s)$	-0.1	1	-4	None
$h_{2,s}$	Output response to monetary shock	-0.3	0.5	-2	None
Priors for structural variances $\mathbf{D}_s \vert \mathbf{A}_s$					
Gamma distribution					
$d_{ii,s}^{-1}$	Reciprocal of variance	$1/\left(\mathbf{a}_{i,s}'\hat{\mathbf{S}}\mathbf{a}_{i,s}\right))$	$1/\left(\sqrt{2}\mathbf{a}_{i,s}'\hat{\mathbf{S}}\mathbf{a}_{i,s}\right)$	–	$d_{ii,s} > 0$
Priors for lag coefficients $\mathbf{B}_s \vert \mathbf{A}_s, \mathbf{D}_s$					
Normal distribution					
$\mathbf{b}_{i,s}$	Lagged coefficients of equation i	$\eta'\mathbf{a}_{i,s}$	$\sqrt{d_{ii,s}\mathbf{M}_i}$	–	None
In addition,					
$b_{33,s}$	Third element of monetary equation	ρ_s	$\sqrt{d_{33,s}/10}$	–	None
Priors for transition probabilities P					
Beta distribution with $\alpha = 1$ and $\beta = 1$					
$p_{s',s}$	Transition probability from s to s'	0.5	0.3	–	$0 \leq p_{s',s} \leq 1$

Discussion of estimation results of the baseline model

After having the MCMC draws, we need a normalization rule to distinguish regime 1 from regime 2 because the unnormalized posterior distribution is by construction bimodal. To that end, Figure 1.1 plots the output gap, inflation, and interest rates with recession shading from the NBER. A closer look reveals that recessions are associated with more volatile interest rates, inflation, and the output gap, thus natural proposals for a normalization rule are to use either the constants, the shock variances, or the transition probabilities.

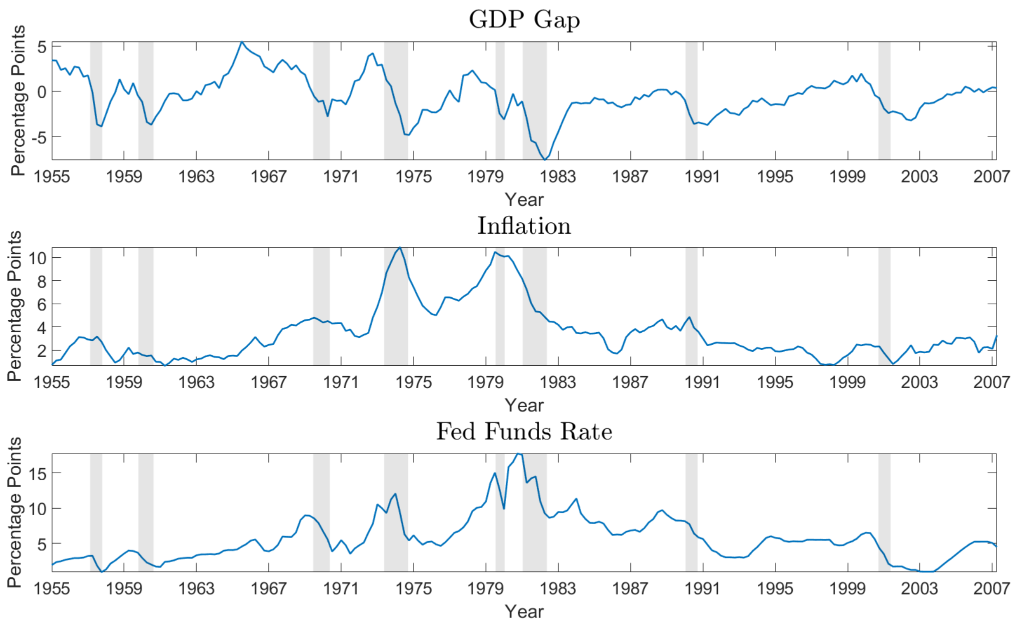

Figure 1.1: **Plots of the output gap, inflation, and fed funds rates**. Output gap is the log difference between real and potential GDP, multiplied by 100. Inflation rate is the log difference between the Y/Y change of the PCE index, multiplied by 100. Fed fund rates is Effective Federal Funds Rate. Shaded area indicates NBER recession periods. More detailed descriptions and data sources are in Appendix A.3.

Figure 1.2 displays some exploratory analyses corresponding to three different normalization rules. The first, second, and third row are exemplary scatter plots of certain model parameters against the (log) variances of monetary shock ($\ln(d_{33})$), the constant of the first structural equation (k^s), and the probability of staying in regime 2 (p_{22}) respectively. Since recessions are associated with lower-than-average GDP growths, one might think that the constant k^s would isolate the two regimes. However, the second row of Figure 1.2 shows that it is not the case. In fact, only the monetary policy shock variances can separate the two regimes, and thus I normalize regime 1 as the regime with the low variance and regime 2 as the one with the high variance. My normalization rule is thus

$$d_{33,1} < d_{33,2}$$

I label the first regime as the ”expansion” regime and the second one as the ”recession” regime. This rule is also easy to apply: after having all the MCMC draws from the unnormalized posterior distribution, we only need to swap parameters between the two states when the constraint is violated.

How well does my model identify actual recessions? Figure 1.3 reports the smoothed probabilities of the ”recession” regime, averaging over all results calculated from each draw of the posterior distribution. Overall, the model captures many recessions, including the beginning of the Great Recession. However, it overestimates the probability of recessions during some periods in the 70s and 80s, while underestimates this probability in the last two recessions before the Great Recession.

How long is a recession or an expansion expected to last? Figure 1.4 plots the prior and posterior distributions for the transition probabilities. The data are highly informative for those parameters even when I use uninformative priors (e.g. uniform distributions over $(0,1)$). The average transition probability from recessions to expansions is about 0.2, whereas the average transition probability from expansions to recessions is about 0.05. Therefore, my model implies that on average, recessions last 5 quarters, whereas expansions last 20 quarters.[18]

[18]Define D as the duration of state i, then from Kim & Nelson (1999), $\mathbb{E}(D) = \frac{1}{1-p_{ii}}$.

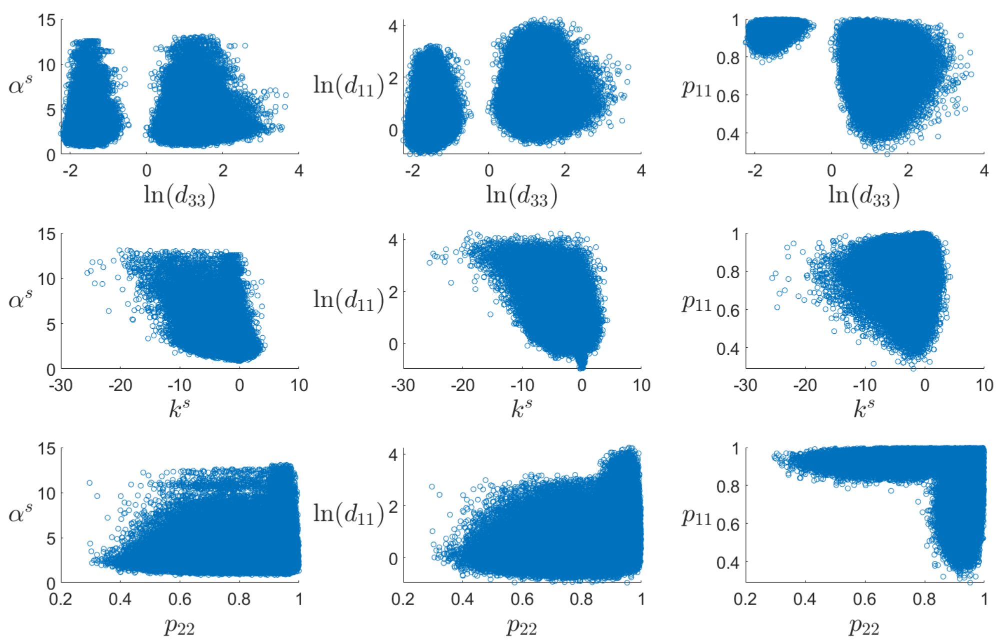

Figure 1.2: **Normalization rules for the baseline model**. This figure shows scatter plots from the unnormalized MCMC output of the baseline model against three candidates for the normalization rules. From top to bottom, the three candidates are the natural log of the monetary shock variance ($\ln(d_{33})$), the constant of the first equation (k^s), and the probability of staying in state 2 (p_{22}). From left to right, the variables on the vertical axes are the effect of π on supply (α^s), the natural log of the supply shock variance ($\ln(d_{11})$), and the probability of staying in state 1 (p_{11}).

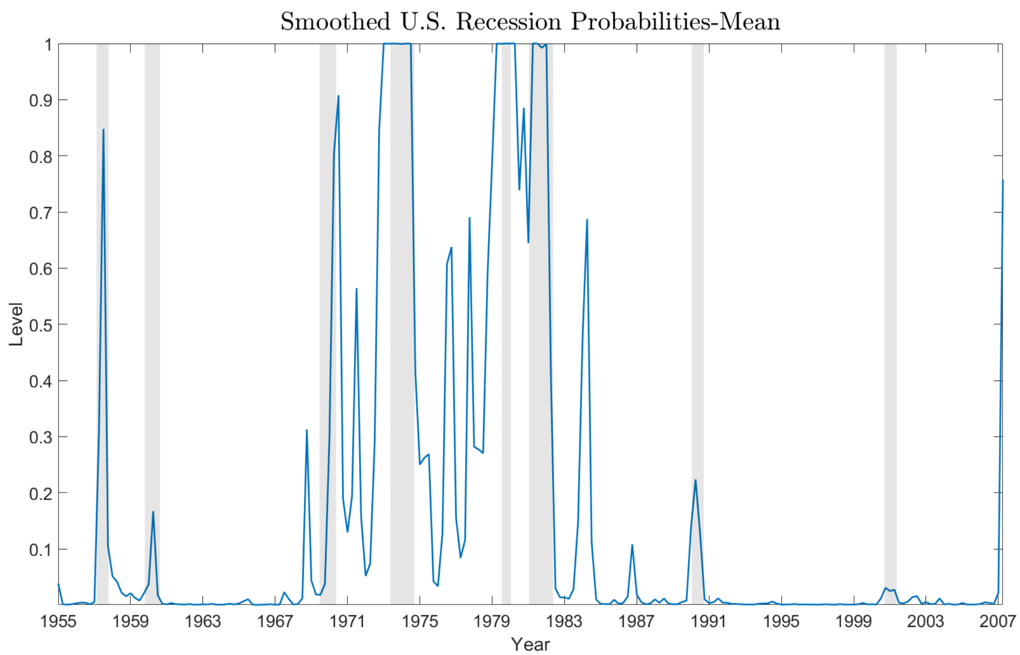

Figure 1.3: **Smoothed probabilities of recession from the baseline model**. This figure shows the smooth probabilities of recessions from the baseline model. The recession regime is defined as the regime that has a larger structural shock variance for the Fed Fund rate. For each draw from the posterior distribution, I calculate the smoothed probabilities by the Hamilton filter, and then I take the mean of the smoothed probabilities across all models. The sample period is 1954Q3 to 2007Q4.

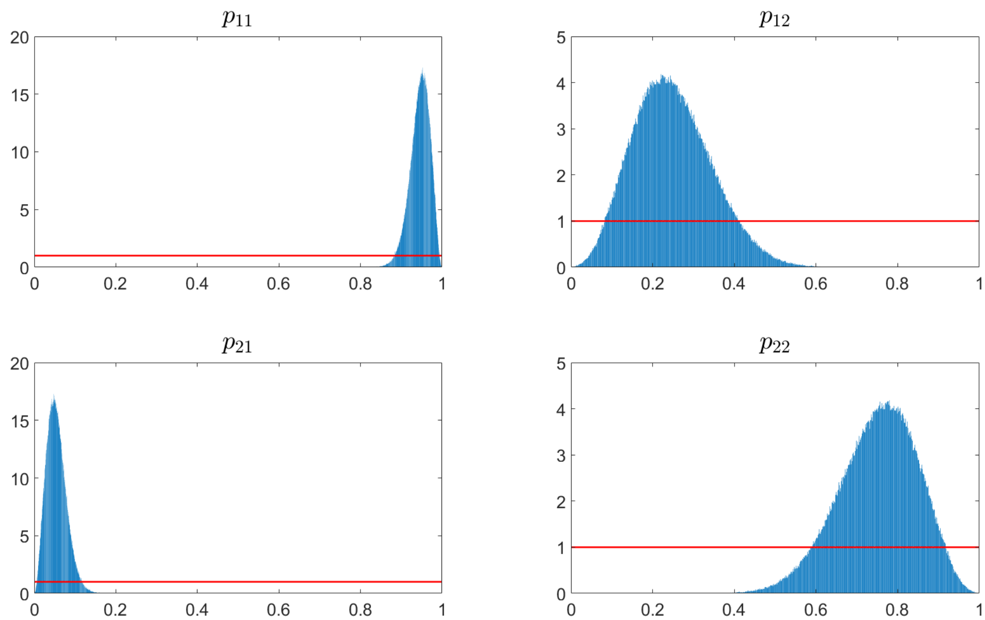

Figure 1.4: **Prior and posterior distributions of transition probabilities of the baseline model**. This figure shows the prior and posterior distributions of the transition probabilities (i.e. elements of the matrix **P**) of the baseline model. The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2007Q4.

Furthermore, the posterior distributions also reveal some differences between parameters across the two regimes. Figure 1.5 highlights the most significant difference between recessions and expansions: the variances of the structural shocks. It also provides further support for the normalization rule. Although the model is normalized by the variance of the monetary policy shock, all six posterior distributions are unimodal with those in recessions exhibit larger magnitudes than those in expansions.

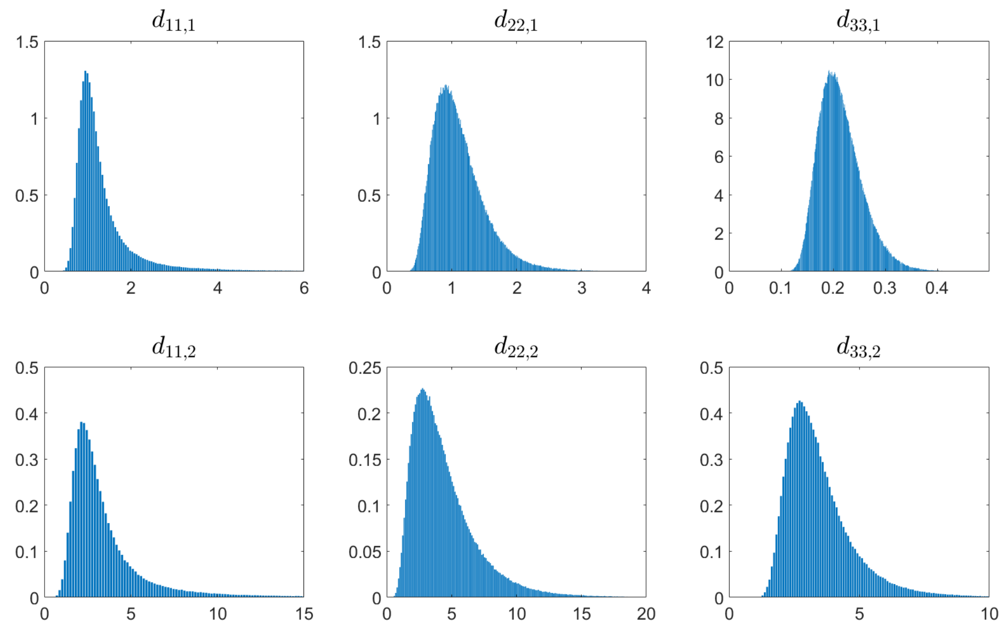

Figure 1.5: **Posterior distributions of the structural shock variances of the baseline model**. This figure shows the posterior distributions of the structural shock variances of the baseline model (i.e. elements of the matrix $\mathbf{D_1}$ and the matrix $\mathbf{D_2}$). The sample period is between 1954Q3 and 2007Q4.

Notwithstanding, the rest of the structural parameters are mostly comparable. Figures 1.6 and 1.7 summarize the prior and posterior distributions for the contemporaneous parameters in recessions and expansions respectively, while Table 1.3 reports the median values of the estimated parameters together with their 95% credible set. In general, the posterior distributions of parameters in the recession regime have a wider spread because there are fewer observations for recessions than expansions during the sample period. The most notable difference is the posterior distribution of ρ, the interest rate smoothing parameter. The one associated with the recession regime has a wider spread and a lower mode value than its counterpart in the expansion regime. This result suggests that monetary policy is less persistent during recessions, perhaps because monetary authorities need to change their policy more abruptly to counter the recessionary effects.

Table 1.3: **Estimation results: baseline model**. This table shows the estimated value for parameters in $\mathbf{A}^s$, $\mathbf{D}^s$, $\mathbf{B}^s$ and $\mathbf{P}$ in the baseline model. It shows the median value of the posterior distribution of the parameters together with their 95% credible sets in parentheses.

Estimated parameters				
Parameter	**Estimated values**	**Parameter**	**Estimated values**	**Meaning**
α_1^s	2.5 (1.6 , 5.8)	α_2^s	2.7 (1.8 , 6.8)	Effect of π on supply
β_1^d	-1.3 (-2.6 , -0.2)	β_2^d	-0.4 (-2.5 , 0.7)	Effect of π on demand
γ_1^d	-0.7 (-1.4 , -0.1)	γ_2^d	-0.5 (-1.2 , -0.04)	Effect of r on demand
ψ_1^y	1.1 (0.7 , 1.9)	ψ_2^y	1.5 (0.5 , 3.8)	Fed response to y
ψ_1^π	1.7 (1.1 , 2.6)	ψ_2^π	1.5 (0.7 , 2.7)	Fed response to π
ρ_1	0.6 (0.5 , 0.8)	ρ_2	0.4 (0.1 , 0.7)	Interest rate smoothing
$d_{11,1}$	1.1 (0.7 , 4.2)	$d_{11,2}$	2.8 (1.3 , 13.4)	Supply shock variance
$d_{22,1}$	1.0 (0.6 , 2.1)	$d_{22,2}$	3.8 (1.4 , 11.2)	Demand shock variance
$d_{33,1}$	0.2 (0.1 , 0.3)	$d_{33,2}$	3.2 (1.8 , 7.1)	Monetary shock variance
p_{11}	0.95 (0.89 , 0.99)	p_{12}	0.24 (0.08 , 0.45)	Transition probabilities
p_{21}	0.05 (0.01 , 0.11)	p_{22}	0.76 (0.55 , 0.92)	Transition probabilities

Due to the similarity of the structural parameters, the SIRFs between the two regimes also show little difference. Figure 1.8 shows the SIRFs associated with recessions and expansions. The median point estimates of the SIRFs are close to each other even though there is much more

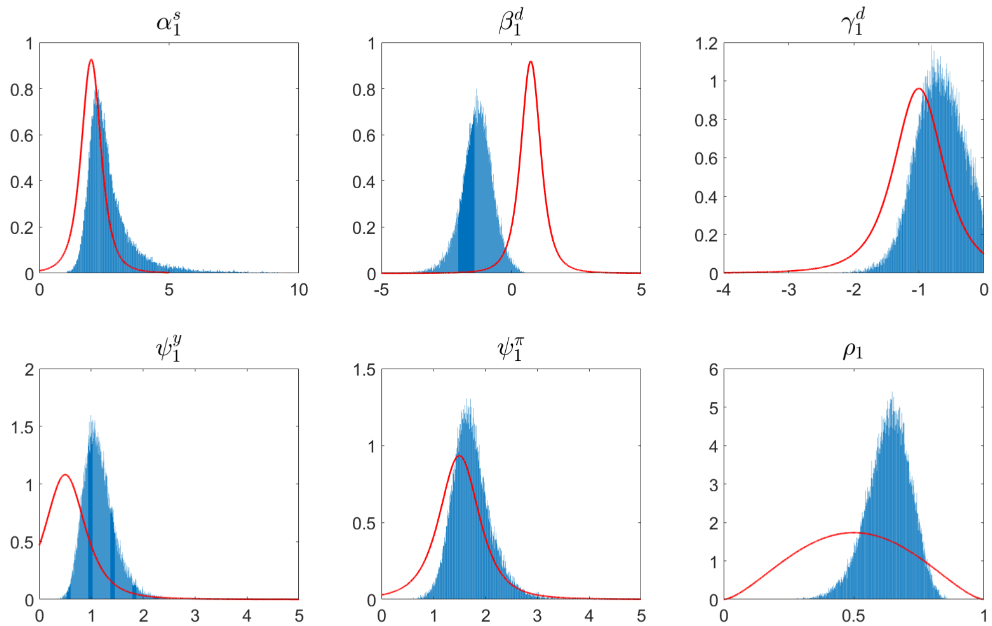

Figure 1.6: **Prior and posterior distributions of contemporaneous structural parameters in state 1 of the baseline model**. This figure shows the prior and posterior distributions of contemporaneous structural parameters (i.e. elements of the matrix $\mathbf{A_1}$) of the baseline model. The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2007Q4.

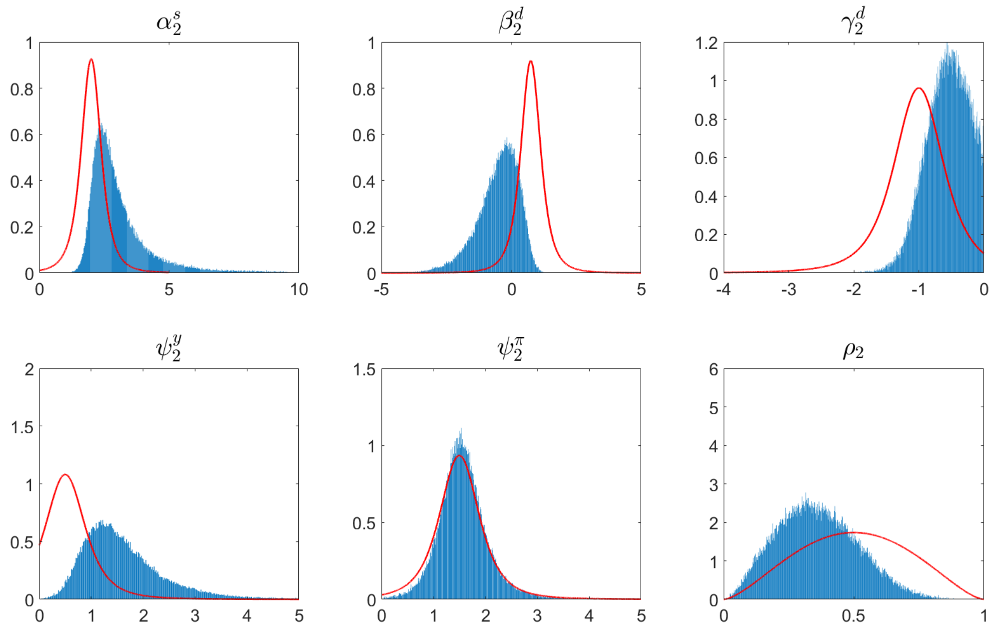

Figure 1.7: **Prior and posterior distributions of contemporaneous structural parameters in state 2 of the baseline model**. This figure shows the prior and posterior distributions of contemporaneous structural parameters (i.e. elements of the matrix $\mathbf{A_2}$) of the baseline model. The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2007Q4.

uncertainty during recessions. In particular, the posterior medians in the third column show that a monetary policy shock that raises the fed funds rate by one percentage point will decrease the output gap by 0.5 percentage points and the inflation rate by 0.2 percentage points upon impact in both regimes.[19] The shock appears to have stronger long-run effects on the inflation rate and less persistent effects on the fed funds rate in recessions, however, those median estimates are subject to considerable uncertainty. Thus, my empirical evidence suggests that the structural coefficients do not change much across regimes, and hence, the propagation of shocks are similar during recessions and expansions. This result stands in contrast with the literature supporting the asymmetric effects of monetary policy.

Bayes factor test for regime changes

Given the estimation result, is the 2-state Markov-switching SVAR an improvement over linear sign-restricted SVAR? To answer that question, I calculate the Bayes factor between the 2-state Markov-switching SVAR, M_2, and the linear sign-restricted SVAR without regime switches, M_1.[20] Table 1.4 shows the marginal likelihood for each model together with the test statistics. For ease of computation and comparison with the frequentist likelihood ratios, I calculate my statistics as two times the log transformation of the Bayes factor. The results are also robust to different computational methods.

How high does the Bayes factor have to be before we decide that model M_2 is a better description of the data than model M_1? Jeffreys (1967) emphasizes that the Bayes factor is a guide to decision making, and thus the meaning of its magnitude depends on the application. In some situations, decision-makers might want to see a very high number for the Bayes factor before they reject the baseline model.[21] Table 1.5 shows a modified version of Jeffrey's guideline proposed

[19] I normalize the SIRFs for the monetary policy shock but I leave the SIRFs for the other two shocks intact. Thus, the first and second column of Figure 1.8 still show the effects of one unit increase of the supply and demand shock respectively.

[20] This model is similar to that of Baumeister & Hamilton (2018), except that my sample period is longer.

[21] "... (The Bayes factor) is not a physical magnitude. Its function is to grade the decisiveness of the evidence. It makes little difference to the null hypothesis whether the odds are 10 to 1 or 100 to 1 against it, and in practice

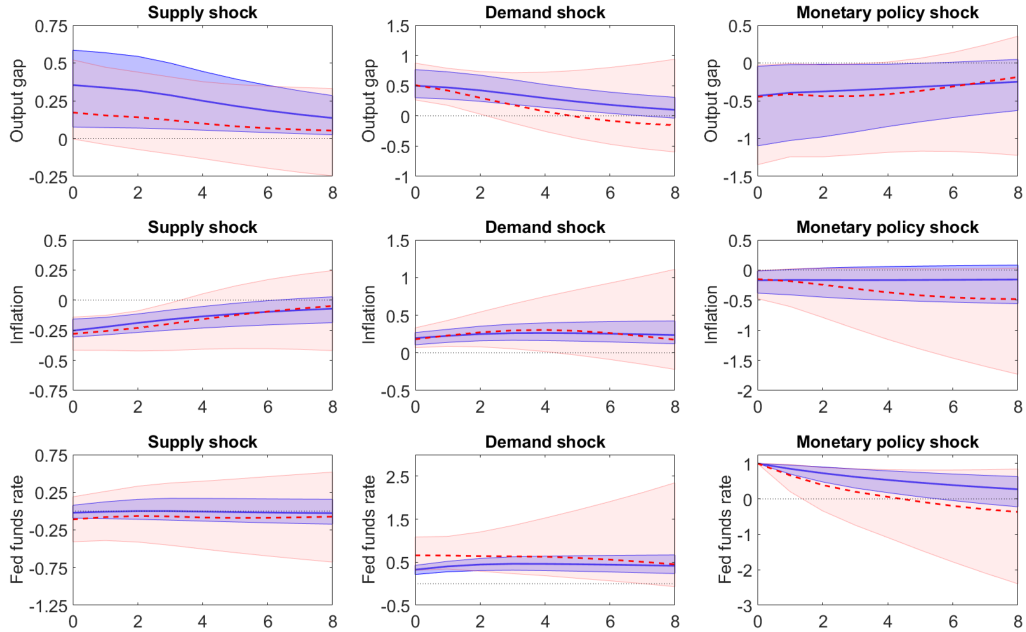

Figure 1.8: **SIRFs comparison between state 1 and state 2: baseline model**. This figure shows the estimated SIRFs of the sign-restricted 2-state MS-SVAR model. The sample period is 1954Q3 to 2007Q4. The solid blue line is the median SIRF of state 1, while the dashed red line is the median SIRF of state 2. The shaded blue region is the 95 percent credible set for state 1, while the shaded red region is the 95 percent credible set for state 2.

Table 1.4: **Bayes factor test of regime change**. This table shows the natural log of the marginal likelihoods for the baseline (M_1) and alternative model (M_2), together with their difference. The alternative model is the sign-restricted 2-state MS-SVAR model, whereas the baseline model is the linear sign-restricted SVAR model without changes in regime. "Geweke" refers to the method proposed by Geweke, 1999, and "SWZ" refers to the one proposed by Sims, Waggoner & Zha, 2008. The parameter τ specifies the level of truncation of the posterior distribution. A value for τ of 0.5 means that 50% of the posterior draws are used, and a value for τ of 0.9 means that 90% of the posterior draws are used.

Test of regime change-Bayes' factor

	$\ln\mathbb{P}(\mathbf{Y}_T\|M_1)$	$\ln\mathbb{P}(\mathbf{Y}_T\|M_2)$	$2\times(\ln\mathbb{P}(\mathbf{Y}_T\|M_2)-\ln\mathbb{P}(\mathbf{Y}_T\|M_1))$
Geweke ($\tau = 0.5$)	-643	-594	98
Geweke ($\tau = 0.9$)	-643	-593	98
SWZ ($\tau = 0.5$)	-644	-597	96
SWZ ($\tau = 0.9$)	-645	-596	98

in Kass & Raftery (1995). It describes different values of the test statistics, their corresponding Bayes factors, and their interpretation.[22] In this case, the test statistic is almost 100. Thus, according to Jeffreys' criteria, there is *very strong* evidence that the 2-state Markov-switching sign-restricted SVAR is a better description of the data than the alternative model.

Table 1.5: **Jeffrey's criteria for model selection**, originally proposed by Jeffreys, 1967. I use the modified version in Kass & Raftery, 1995. The Bayes factor, B_{21}, is the ratio of the marginal likelihood of the alternative model (i.e. M_2) to the marginal likelihood of the baseline model (i.e. M_1). When the prior belief is that the probabilities of the two models are equal, the Bayes factor is also the odds ratio between the alternative and baseline models.

Jeffrey's criteria		
2log($\mathbf{B}_{21}$)	$\mathbf{B}_{21}$	**Evidence against $\mathbf{M}_1$**
0 to 2	1 to 3	Not worth more than a bare mentioning
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

1.3.3 A regime-switching sign-restricted SVAR model of monetary policy augmented with NBER recession indicator

Because the smoothed probabilities in Figure 1.3 do not coincide precisely with NBER recessions and expansions, some may be concerned about my interpretation of the two regimes. As a robustness check, I augment the baseline model by NBER recession dates as an additional indicator. Here I discuss my priors for the quality of the indicator and present the estimation results.

no difference at all whether they are 10^4 or 10^{10} to 1 against it. In any case whatever alternative is most strongly supported will be set up as the hypothesis for use until further notice." *(Jeffreys, 1967, Appendix B)*

[22]Simulations from Efron, Gous, Kass, Datta & Lahiri (2001) suggests that Jeffrey' scale generally are more conservative than the frequentist p-value.

Priors for the quality of the NBER recession indicator

The addition of the NBER recession indicator adds four new parameters corresponding to elements of the matrix **G**. Let Z denote the NBER indicator, the regime indicator matrix is

$$\mathbf{G} = \begin{bmatrix} P(Z_t = 1|S_t = 1) & P(Z_t = 1|S_t = 2) \\ P(Z_t = 2|S_t = 1) & P(Z_t = 2|S_t = 2) \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

Recall that each element of the matrix **G** represents the information in Z about the regime. Since I don't want to exert any strong belief about the quality of information, I use the Dirichlet(1,1) distribution for each column of that matrix. It is equivalent to use a uniform distribution over $(0,1)$ as the prior for each element. As in the general formulation, I assume independent priors across columns. Formally, the priors are

$$P(\mathbf{G}) = P(g_{11}, g_{21}) P(g_{12}, g_{22})$$
$$(g_{11}, g_{21}) \sim D(1,1)$$
$$(g_{12}, g_{22}) \sim D(1,1)$$

Table 1.6 shows the prior distributions of the baseline model augmented with the NBER recession indicator. It is similar to that of the baseline model, Table 1.2, except now we have four additional parameters from the regime indicator matrix, **G**.

Discussion of estimation results of the model using NBER recession indicator

I use the same normalization rule as that of the baseline model (that is, $d_{33,1} < d_{33,2}$) and label regime 1 as expansion and regime 2 as recession. Figure 1.9 shows the prior and posterior distributions for elements of matrix **G**. Even with uniform priors, the NBER indicator is valuable in separating the two regimes. The posterior of g_{11} notably concentrates closer to 1 than that of

Table 1.6: **Prior distributions of the model with the NBER recession indicator**. This table shows the prior distribution of $\mathbf{A}_s$, $\mathbf{D}_s$, $\mathbf{B}_s$, $\mathbf{P}$, and $\mathbf{G}$, together with their hyper-parameters. For Student t and Asymmetric t distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

Parameter	Meaning	Location	Scale	Skew	Sign restriction
	Priors for the contemporaneous coefficients of $\mathbf{A}_s$ and elements of $\mathbf{A}_s^{-1}$				
	Student t distribution with 3 degrees of freedom				
α_s^s	Effect of π on supply	2	0.4	–	$\alpha_s^s \geq 0$
β_s^d	Effect of π on demand	0.75	0.4	–	None
γ_s^d	Effect of r on demand	-1	0.4	–	$\gamma_s^d \leq 0$
ψ_s^y	Fed response to y	0.5	0.4	–	$\psi_s^y \geq 0$
ψ_s^π	Fed response to π	1.5	0.4	–	$\psi_s^\pi \geq 0$
	Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$				
ρ_s	Interest rate smoothing	0.5	0.2	–	$0 \leq \rho_s \leq 1$
	Asymmetric t distribution with 3 degrees of freedom				
$h_{1,s}$	Part of $\det(\mathbf{A}_s)$	-0.1	1	-4	None
$h_{2,s}$	Output response to monetary shock	-0.3	0.5	-2	None
	Priors for the variances of structural shocks $\mathbf{D}_s \vert \mathbf{A}_s$				
	Gamma distribution				
$d_{ii,s}^{-1}$	Reciprocal of variance	$1/\left(\mathbf{a}_{i,s}'\hat{\mathbf{S}}\mathbf{a}_{i,s}\right))$	$1/\left(\sqrt{2}\mathbf{a}_{i,s}'\hat{\mathbf{S}}\mathbf{a}_{i,s}\right)$	–	$d_{ii,s} > 0$
	Priors for the lag coefficients $\mathbf{B}_s \vert \mathbf{A}_s, \mathbf{D}_s$				
	Normal distribution				
$\mathbf{b}_{i,s}$	Lagged coefficients of equation i	$\eta'\mathbf{a}_{i,s}$	$\sqrt{d_{ii,s}\mathbf{M}_i}$	–	None
In addition,					
$b_{33,s}$	Third element of monetary equation	ρ_s	$\sqrt{d_{33,s}/10}$	–	None
	Priors for the transition probabilities P				
	Beta distribution with $\alpha = 1$ and $\beta = 1$				
$p_{s',s}$	Transition probability from s to s'	0.5	0.3	–	$0 \leq p_{s',s} \leq 1$
	Priors for the regime indicator matrix G				
	Beta distribution with $\alpha = 1$ and $\beta = 1$				
$g_{s',s}$	Probability that the indicator is s' when the state is s	0.5	0.3	–	$0 \leq g_{s',s} \leq 1$

g_{22} because there are more observations in expansions than in recessions. The most significant difference from the baseline model is seen in Figure 1.10, which shows the smoothed probabilities of recession (i.e. regime 2); almost every peak of the smoothed probabilities now coincides with the recession shading from the NBER.

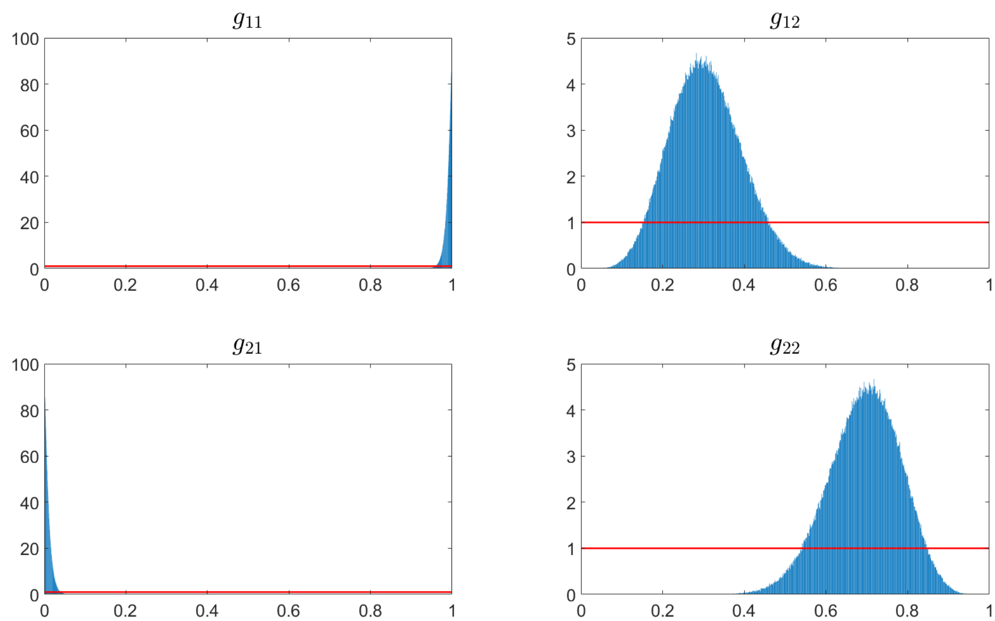

Figure 1.9: **Prior and posterior distributions of elements of G**. This figure shows the prior and posterior distributions of elements of the regime indicator matrix, **G**. The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2007Q4.

Is the NBER indicator informative about the two regimes? To answer that question, I estimate another model where the NBER indicator is restricted to be uninformative. Technically, I restrict the columns of the matrix **G** to be the same as follows

$$\mathbf{G} = \begin{bmatrix} P(Z_t = 1|S_t = 1) & P(Z_t = 1|S_t = 2) \\ P(Z_t = 2|S_t = 1) & P(Z_t = 2|S_t = 2) \end{bmatrix} = \begin{bmatrix} g & g \\ 1-g & 1-g \end{bmatrix}$$

The Bayes factor between the unrestricted model and the restricted model will tell us whether the indicator is informative. Table 1.7 shows the Bayes factor test for the informativeness of the regime indicator. According to the test statistics and the Jeffrey's criteria, there is very strong evidence that the NBER indicator is informative.

Table 1.7: **Bayes factor test of the informativeness of the regime indicator**. This table shows the natural log of the marginal likelihoods for the baseline (M_1) and alternative model (M_2), together with their difference. The alternative model is the sign-restricted 2-state MS-SVAR model with the regime indicator, whereas the baseline model is the sign-restrictied 2-state MS-SVAR model where the regime indicator is restricted to be uninformative. "Geweke" refers to the method proposed by Geweke, 1999, and "SWZ" refers to the one proposed by Sims, Waggoner & Zha, 2008. The parameter τ specifies the level of truncation of the posterior distribution. A value for τ of 0.5 means that 50% of the posterior draws are used, and a value for τ of 0.9 means that 90% of the posterior draws are used.

Test of informativeness of the regime indicator-Bayes' factor			
	$\ln\mathbb{P}(\mathbf{W}_T\|M_1)$	$\ln\mathbb{P}(\mathbf{W}_T\|M_2)$	$2\times(\ln\mathbb{P}(\mathbf{W}_T\|M_2)-\ln\mathbb{P}(\mathbf{W}_T\|M_1))$
Geweke ($\tau=0.5$)	-677	-649	57
Geweke ($\tau=0.9$)	-677	-648	57
SWZ ($\tau=0.5$)	-680	-652	57
SWZ ($\tau=0.9$)	-679	-651	56

Lastly, Table 1.8 summarizes the new estimation results. Despite better identification of the two regimes, the estimated parameters are versy similar to those of the baseline model. The resulting SIRFs, as shown in Figure 1.11, are comparable across the two regimes. Thus, even when the model is augmented with the NBER indicator, monetary policy shocks still have symmetric effects during recessions and expansions.

Table 1.8: **Estimation results: model with regime indicator**. This table shows the estimated value for parameters in $\mathbf{A}^s$, $\mathbf{D}^s$, $\mathbf{B}^s$, $\mathbf{P}$, and $\mathbf{G}$ in the model with regime indicator. It shows the median value of the posterior distribution of the parameters together with their 95% credible sets in parentheses.

Estimated parameters				
Parameter	**Estimated values**	**Parameter**	**Estimated values**	**Meaning**
α_1^s	2.4	α_2^s	2.6	Effect of π on supply
	(1.6 , 4.7)		(1.7 , 5.3)	
β_1^d	-1.0	β_2^d	-0.4	Effect of π on demand
	(-2 , -0.2)		(-2.1 , 0.7)	
γ_1^d	-0.7	γ_2^d	-0.5	Effect of r on demand
	(-1.4 , -0.1)		(-1.3 , -0.04)	
ψ_1^y	1.1	ψ_2^y	1.5	Fed response to y
	(0.6 , 2.1)		(0.5 , 3.8)	
ψ_1^π	1.6	ψ_2^π	1.6	Fed response to π
	(1 , 2.5)		(0.7 , 2.6)	
ρ_1	0.7	ρ_2	0.3	Interest rate smoothing
	(0.5 , 0.8)		(0.1 , 0.7)	
$d_{11,1}$	1.1	$d_{11,2}$	2.6	Supply shock variance
	(0.6 , 3.2)		(1.3 , 8.8)	
$d_{22,1}$	0.9	$d_{22,2}$	3.3	Demand shock variance
	(0.5 , 1.7)		(1.3 , 9.4)	
$d_{33,1}$	0.2	$d_{33,2}$	2.8	Monetary shock variance
	(0.2 , 0.3)		(1.7 , 5.9)	
p_{11}	0.95	p_{12}	0.23	Transition probabilities
	(0.90 , 0.98)		(0.11 , 0.39)	
p_{21}	0.05	p_{22}	0.77	Transition probabilities
	(0.02 , 0.10)		(0.61 , 0.89)	
g_{11}	0.99	g_{12}	0.30	Elements of regime indicator matrix
	(0.97 , 1.00)		(0.15 , 0.49)	
g_{21}	0.01	g_{22}	0.70	Elements of regime indicator matrix
	(0.00 , 0.03)		(0.51 , 0.85)	

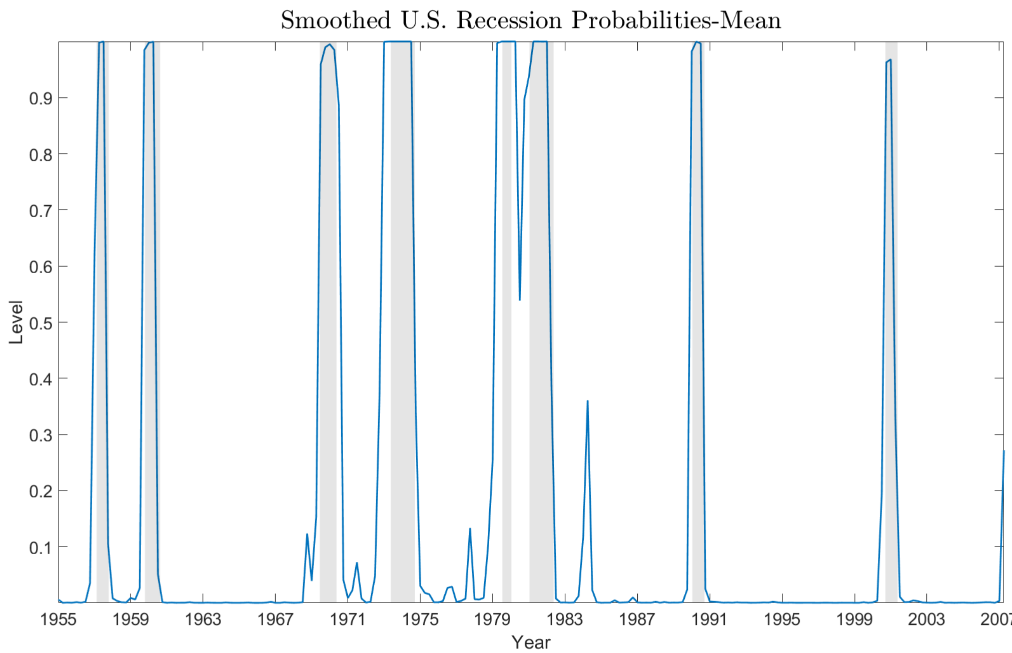

Figure 1.10: **Smoothed probabilities of recession from the model with NBER recession indicator**. This figure shows the the smooth probabilities of recessions from the baseline model augmented with the NBER recession indicator. The recession regime is defined as the regime that has a larger structural shock variance for the Fed Fund rate. For each draw from the posterior distribution, I calculate the smoothed probabilities by the Hamilton filter, and then I take the mean of the smoothed probabilities across all models. The sample period is 1954Q3 to 2007Q4.

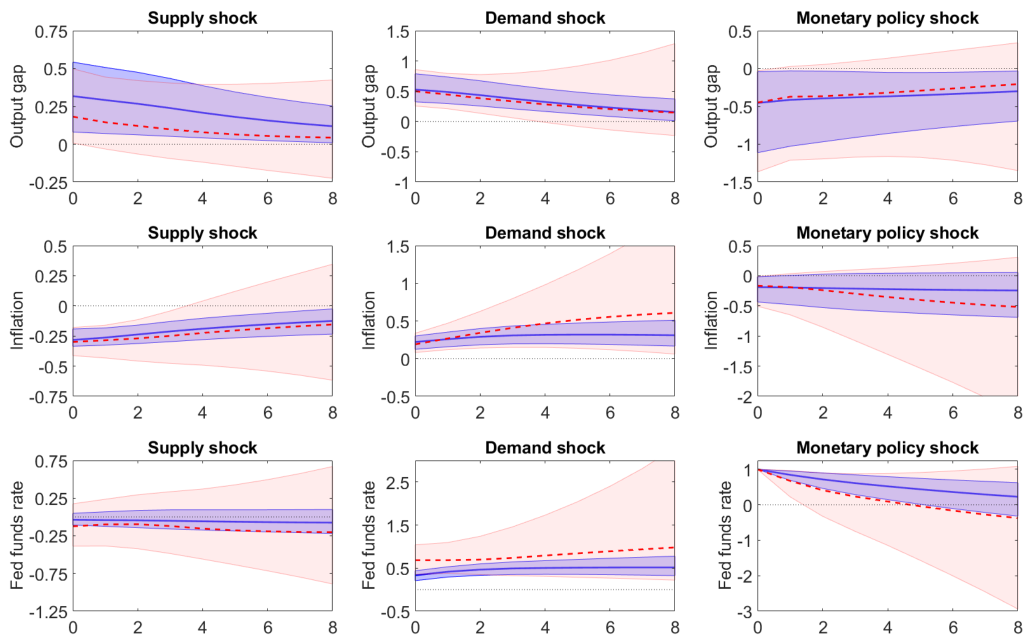

Figure 1.11: **SIRFs comparison between state 1 and state 2: model with NBER recession indicator**. This figure shows the estimated SIRFs of the sign-restricted 2-state MS-SVAR model augmented with the NBER recession indicator. The sample period is between 1954Q3 and 2007Q4. The solid blue line is the median SIRF of state 1, while the dashed red line is the median SIRF of state 2. The shaded blue region is the 95 percent credible set for state 1, while the shaded red region is the 95 percent credible set for state 2.

1.3.4 A regime-switching sign-restricted SVAR model of monetary policy augmented with NBER recession indicator and asymmetric priors

Here I explore whether the conclusion would change if we estimate the previous model with asymmetric priors, starting with a strong prior expectation that monetary policy is stronger in one regime. This specification uses both the NBER recession indicator to help identify the two states and a prior belief that the economic relationship and mechanism of the two states are different.

Asymmetric priors for the contemporaneous parameters

The literature points out several reasons why conventional monetary policy might produce asymmetric effects due to structural changes in the economy during recessions and expansions. One dominant channel is via the change in the intertemporal elasticity of substitution (IES): a higher IES implies stronger effects of monetary policy. This channel will be incorporated as priors on the contemporaneous parameters of the two states. In particular, I use asymmetric priors on the responses of output demand to inflation $\left(\beta_1^d, \beta_2^d\right)$ and the responses of output demand to interest rate $\left(\gamma_1^d, \gamma_2^d\right)$. I set the modes of β_2^d and γ_2^d to be twice as large as their counterparts in absolute magnitude while leaving the rest of the priors the same as the previous specification with the NBER recession indicator.[23]

Table 1.9 summarizes the asymmetric prior distributions. Intuitively, the priors are that in one regime, output demand responds more to the interest rate and inflation. In that regime, a monetary policy shock that raises the fed funds rate by one percentage point will, on average, decrease the output gap by 1.5 percentage points and the inflation rate by 0.7 percentage points upon impact. In the other regime, a similar shock will lower both the output gap and inflation by 0.8 percentage points and 0.4 percentage points respectively.

[23] Baumeister & Hamilton (2018) use 0.5 for the IES to derive the prior modes for β^d and γ^d. Under the same assumptions, doubling the modes of β^d and γ^d is equivalent to using 1 for the IES.

Table 1.9: **Prior distributions of the model with NBER recession indicator and asymmetric priors**. This table shows the prior distributions of $\mathbf{A}_s$ together with their hyper-parameters. For Student t distribution, the location parameter refers to the mode. For Beta distribution, the location parameter is the mean and the scale parameter is the standard deviation. The prior distributions of $\mathbf{A}_s^{-1}$, $\mathbf{D}_s$, $\mathbf{B}_s$, $\mathbf{P}$, and $\mathbf{G}$ are similar as those of the model augmented with NBER recession indicator.

Parameter	Meaning	Location	Scale	Skew	Sign restriction
Asymmetric priors for contemporaneous coefficients, $p_1(.)$					
Student t distribution with 3 degrees of freedom					
β_s^d	Effect of π on demand	0.75	0.4	–	None
γ_s^d	Effect of r on demand	-1	0.4	–	$\gamma_s^d \leq 0$
Asymmetric priors for contemporaneous coefficients, $p_2(.)$					
Student-t distribution with 3 degrees of freedom					
β_s^d	Effect of π on demand	1.5	0.4	–	None
γ_s^d	Effect of r on demand	-2	0.4	–	$\gamma_s^d \leq 0$
Symmetric priors for contemporaneous coefficients					
Student t distribution with 3 degrees of freedom					
α_s^s	Effect of π on supply	2	0.4	–	$\alpha_s^s \geq 0$
ψ_s^y	Fed response to y	0.5	0.4	–	$\psi_s^y \geq 0$
ψ_s^π	Fed response to π	1.5	0.4	–	$\psi_s^\pi \geq 0$
Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$					
ρ_s	Interest rate smoothing	0.5	0.2	–	$0 \leq \rho_s \leq 1$

A straightforward application of the permutation sampler no longer works because the posterior distribution is now asymmetric even though it can still be multimodal. A simple way to accommodate such situations is to use a mixture prior as recommended by Kaufmann & Frühwirth-Schnatter (2002). Specifically, the following mixture prior allows us to express the prior belief that monetary policy exerts stronger effects in one regime without stating which regime this will be.

$$P(\mathbf{A}_s) = \frac{1}{2}\sum_{u=1}^{2} p(\alpha_s^s)\, p_u\left(\beta_s^d\right) p_u\left(\gamma_s^d\right) p(\psi_s^y)\, p(\psi_s^\pi)\, p(\rho_s)\, p(h_{1,s})^{\zeta_{h_{1,s}}}\, p(h_{2,s})^{\zeta_{h_{2,s}}} \quad \text{for } s = 1,2 \tag{1.44}$$

Because this mixture prior is still symmetric, the estimation procedure remains the same.

Discussion of estimation results of the model using NBER indicator and asymmetric priors

I use the same normalization rule as that of the baseline model (that is, $d_{33,1} < d_{33,2}$) and label regime 1 as expansion and regime 2 as recession. Table 1.10 shows the estimation results, while Figure 1.12 shows the corresponding SIRFs for the two regimes. Despite the asymmetric priors that imply monetary policy to be more powerful in one regime, both the point estimates of the structural parameters and the SIRFs are essentially the same to that of the model with symmetric priors. Thus, the evidence in the data that monetary policy has similar effects during recessions and expansions is strong enough to override a strong prior belief that monetary policy has asymmetric effects.

Table 1.10: **Estimation results of the model with NBER recession indicator and asymmetric priors**. This table shows the estimated value for parameters in $\mathbf{A}^s$, $\mathbf{D}^s$, $\mathbf{B}^s$, $\mathbf{P}$, and $\mathbf{G}$ in the model with regime indicator and asymmetric priors. It shows the median value of the posterior distribution of the parameters together with their 95% credible sets in parentheses.

Estimated parameters				
Parameter	**Estimated values**	**Parameter**	**Estimated values**	**Meaning**
α_1^s	2.3	α_2^s	2.6	Effect of π on supply
	(1.6 , 4.5)		(1.7 , 5.1)	
β_1^d	-1.0	β_2^d	-0.4	Effect of π on demand
	(-2.0 , -0.2)		(-2.1 , 0.7)	
γ_1^d	-0.7	γ_2^d	-0.5	Effect of r on demand
	(-1.5 , -0.1)		(-1.3 , -0.04)	
ψ_1^y	1.1	ψ_2^y	1.5	Fed response to y
	(0.6 , 2.0)		(0.5 , 3.7)	
ψ_1^π	1.6	ψ_2^π	1.6	Fed response to π
	(1.0 , 2.5)		(0.8 , 2.6)	
ρ_1	0.7	ρ_2	0.3	Interest rate smoothing
	(0.4 , 0.8)		(0.1 , 0.7)	
$d_{11,1}$	1.1	$d_{11,2}$	2.6	Supply shock variance
	(0.6 , 3.0)		(1.3 , 8.5)	
$d_{22,1}$	0.9	$d_{22,2}$	3.3	Demand shock variance
	(0.5 , 1.8)		(1.3 , 9.6)	
$d_{33,1}$	0.24	$d_{33,2}$	2.8	Monetary shock variance
	(0.17 , 0.36)		(1.7 , 5.9)	
p_{11}	0.95	p_{12}	0.23	Transition probabilities
	(0.90 , 0.98)		(0.11 , 0.39)	
p_{21}	0.05	p_{22}	0.77	Transition probabilities
	(0.02 , 0.10)		(0.61 , 0.89)	
g_{11}	0.99	g_{12}	0.30	Elements of regime indicator matrix
	(0.97 , 1.00)		(0.15 , 0.49)	
g_{21}	0.01	g_{22}	0.70	Elements of regime indicator matrix
	(0.00 , 0.03)		(0.51 , 0.85)	

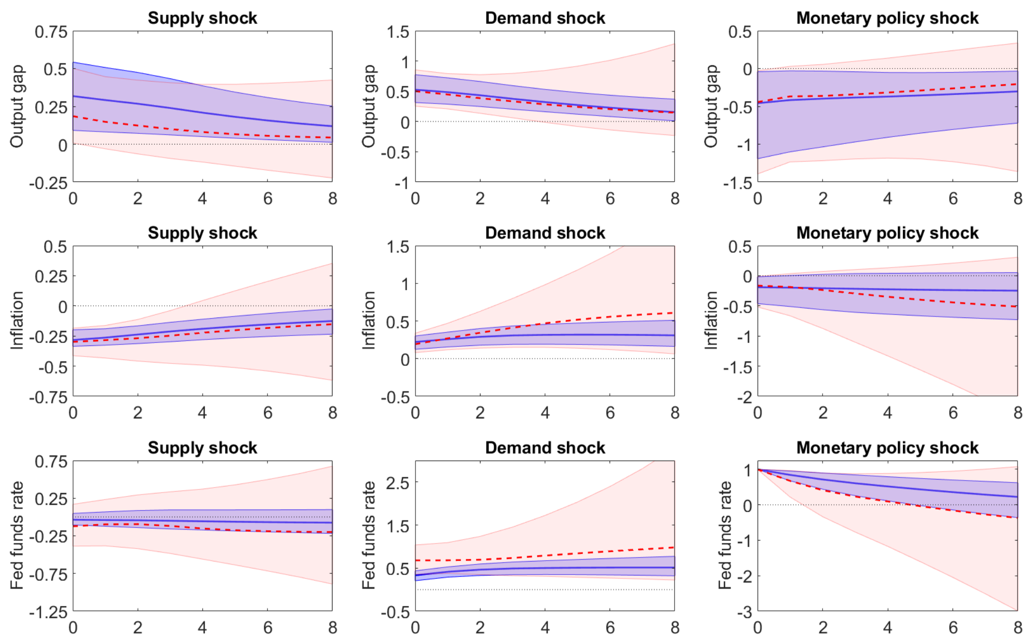

Figure 1.12: **SIRFs comparison between state 1 and state 2: model with NBER recession indicator and asymmetric priors**. This figure shows the estimated SIRFs of the sign-restricted 2-state MS-SVAR model with asymmetric prior. The sample period is 1954Q3 to 2007Q4. The solid blue line is the median SIRF of state 1, while the dashed red line is the median SIRF of state 2. The shaded blue region is the 95 percent credible set for state 1, while the shaded red region is the 95 percent credible set for state 2.

1.4 Conclusion

Macroeconomists are increasingly using more robust methods to identify and estimate dynamic causal effects. Identification by sign restrictions is a promising method emerging from this research agenda. My paper contributes to this rising literature by developing a novel MCMC algorithm to estimate regime-switching SVAR models identified by sign restrictions. This approach offers three improvements over current practices. First, it can impose both sign restrictions and conventional identification strategies. Second, it incorporates priors about both structural parameters and hidden states explicitly without relying on draws from the uniform Haar distribution to rotate the reduced-form VAR, a practice known to impose implicit priors on the dynamic causal effects. Lastly, it accounts for the label-switching problem of Bayesian estimation, which has largely been ignored in the previous econometric literature.

In the empirical application, I revisited the literature on asymmetric effects of conventional monetary policy over the business cycle by estimating a sign-restricted 2-state Markov-switching SVAR model of the output gap, inflation, and interest rate. Although the analysis reveals that the Markov-switching SVAR is a better description of the data than the linear one, the major difference between the two regimes appears to be the variances of their structural shocks instead of other structural parameters. My results suggest that previous work exaggerated changes in coefficients because it neglected changes in variances of structural shocks. I find that monetary policy tightening reduces output and inflation by similar magnitudes in both recessions and expansions. My findings are also robust when the model is augmented with the NBER indicator and asymmetric priors.

Acknowledgements

Chapter 1, in full, is currently being prepared for submission for publication of the material. Lam Nguyen. The dissertation author was the primary author of this chapter.

Chapter 2

Bayesian Inference in Structural Vector Autoregression with Sign Restrictions and External Instruments

Abstract

Instrument validity cannot be tested in a just-identified model, and it is not clear what conclusion to draw when instrument validity is rejected in an over-identified model. In practice researchers tend to regard instruments as valid when they lead to sensible inferences. This paper uses Bayesian methods to formalize this idea. I develop a proxy structural vector autoregression in which prior information from both theory and the empirical literature is incorporated about signs and magnitudes of certain parameters and equilibrium impacts. I use my method to investigate the relevance and validity of three popular instruments for monetary policy shocks, developed by Romer and Romer (2004), Sims and Zha (2006), and Smets and Wouters (2007). I find that all of them are strongly relevant but only that of Smets and Wouters is valid. Furthermore, the empirical analysis demonstrates that my framework can combine information from a relevant and valid instrument with prior information about sign restrictions to improve inference about

structural impulse-response functions.

2.1 Introduction

Starting with Sims (1980), empirical macroeconomists have been developing new robust identification strategies to perform dynamic causal inference in structural vector autoregressions (SVARs). As conventional identifying assumptions have come under intense scrutiny (Ramey (2016), Nakamura & Steinsson (2018)), *identification by instrumental variables* has emerged as a more credible method for causal analysis. The method, also known as Proxy SVAR, was popularized by Stock & Watson (2012a) and Stock & Watson (2018), Mertens & Ravn (2013), Gertler & Karadi (2015), and has been applied to a wide variety of settings.

Is Proxy SVAR really more credible? Its central premise is that imperfect measures can be used as instruments for the true shocks.[1] The key identifying assumptions are that the instruments are correlated with the true shocks (relevance) and uncorrelated with other structural shocks (validity).[2] Although the relevance assumption can be formally tested (Stock & Watson (2018)), the validity assumption is often defended by *ad hoc* arguments. Stock & Watson (2012a) provide suggestive evidence showing that this assumption might fail to hold for the majority of instruments in macroeconomics. However, without further information, the data cannot tell us whether this assumption holds for any particular instrument.

How, then, can researchers select a good instrument among many potentially invalid instruments? I propose to evaluate both instrument relevance and validity by using a *subjective Bayesian* approach that incorporates information from both theory and the empirical literature in

[1]Examples of such instruments are Hamilton (2003), Kilian (2008), Arezki, Ramey & Sheng (2017) (oil shocks) , Romer & Romer (2010), Ramey (2011), Fisher & Peters (2010) (fiscal policy shocks), and Gürkaynak, Sack & Swansonc (2005), Romer & Romer (2004) (monetary policy shocks).

[2]In this paper, I use the terminology in Hamilton (1994b). In some other texts, such as Stock & Watson (2012b), the authors define a valid instrument as an instrument that is relevant and exogenous. Their definition of instrument relevance is the same as mine, and their definition of instrument exogeneity is equivalent to my definition of instrument validity.

the form of sign restrictions. The literature on sign restrictions are pioneered by Uhlig (2005), Faust (1998), and Canova & De Nicolo (2002), and it has increasingly gained popularity with many recent contributions, including Rubio-Ramírez, Waggoner & Zha (2010), Arias, Rubio-Ramírez & Waggoner (2018b), Arias, Caldara & Rubio-Ramírez (2019), Antolín-Díaz & Rubio-Ramírez (2018), and Baumeister & Hamilton (2015), Baumeister & Hamilton (2018), and Baumeister & Hamilton (2019). The surge in popularity is due to the fact that sign restrictions are more consistent with economic theory and thus more robust to model misspecification. My paper suggests using the identifying assumptions from sign restrictions to evaluate the underlying assumption of the instruments.

To implement the idea, I generalize the SVAR framework of Baumeister & Hamilton (2015) to incorporate both sign restrictions and external instruments. Investigation of the validity and relevance assumptions will then be framed as a model selection problem. In particular, instrument validity is investigated by a Bayesian model comparison between a model where the validity assumption holds and one where it does not. If the data favor the first model, the instruments are judged to be valid, otherwise, they are invalid. Similarly, the method can be used to evaluate the relevance assumption.

I apply the technique to analyze the relevance and validity of three monetary policy instruments, proposed by Romer & Romer (2004), Sims & Zha (2006), and Smets & Wouters (2007).[3] I first use various tests of overidentifying restrictions in a frequentist Proxy SVAR framework to demonstrate that at least some of the instruments are invalid and to illustrate the inability of existing methods to differentiate valid instruments from invalid ones. Then, I use the sign-restricted SVAR proposed by Baumeister & Hamilton (2018) as a benchmark to evaluate the underlying assumptions of those instruments. My main finding is that all three instruments are strongly relevant, but only the Smets-Wouters instrument is valid. The reason is that the Romer-

[3]The instrument of Romer & Romer (2004) is constructed by the narrative method, that of Sims & Zha (2006) is estimated from a Markov-switching SVAR model, and that of Smets & Wouters (2007) comes from a medium-scale DSGE model.

Romer and Sims-Zha instruments are highly correlated with both monetary and demand shocks identified from the sign-restricted SVAR. This finding sheds light on Stock & Watson (2012a)'s observation that monetary and fiscal policy shocks, identified by their respective instruments, are highly correlated with each other.

Furthermore, the empirical application establishes that a Bayesian framework can use instruments to identify structural shocks just as in conventional approaches. I first use Romer & Romer (2004)'s specification to study the dynamics of macro aggregates in response to a monetary shock. Then I show that the sign-restricted SVAR augmented with the instrument produces structural impulse-response functions (SIRFs) with similar patterns. In particular, I find that both the output gap and inflation have hump-shaped responses following a monetary policy shock. This finding is consistent with those from previous literature as surveyed in Ramey (2016).

My paper is the first to evaluate the instrument relevance and validity assumptions by sign restrictions, and in doing so, it addresses three strands of literature. First is the econometric literature on SVAR identified by sign restrictions and instrumental variables. The major distinction between my framework and others is that mine transparently and flexibly incorporates *both sign restrictions and instrumental variables*. In contrast to Arias, Rubio-Ramírez & Waggoner (2018a), Ludvigson, Ma & Ng (2017), and Braun & Brüggemann (2017), I use explicit priors and sign restrictions to estimate the SVAR parameters directly without using draws of the rotation matrix to convert the reduced-form VAR, a practice known to impose implicit priors on the dynamic causal effects (Baumeister & Hamilton (2015), Wolf (2018), Watson (2019)). Moreover, my method can easily use multiple proxies to identify a single shock, a common objective which cannot be achieved using Arias, Rubio-Ramírez & Waggoner (2018a)'s algorithm. Compared to the robust Bayesian method of Giacomini, Kitagawa & Read (2019), my parameterization makes better use of information from the previous literature because the SVAR parameters have economic interpretation. Relative to other approaches in Bayesian Proxy SVAR, my method not only enjoys all the advantages over the frequentist counterpart as discussed in Caldara & Herbst

(2019), Bahaj (2014), and Drautzburg (2016) but it also use sign restrictions and informative priors on various model quantities to improve causal inference.

Second, my paper contributes to the econometric literature on instrumental variable regression with imperfect instruments. Early work, such as Conley, Hansen & Rossi (2012), Nevo & Rosen (2012), and Chan & Tobias (2015), relaxed the validity assumption by allowing the instruments to have a small, direct effect on the dependent variable. Although those papers acknowledge that the instrument validity is likely violated, none of them attempts to formally investigate that assumption.[4] More recently, Ludvigson, Ma & Ng (2017) and Braun & Brüggemann (2017) consider sign-restricted SVAR models with external instruments where they restrict the correlations between the structural shocks and the instruments. Because of independent evidence that many instruments are likely invalid, using the instruments to restrict the model could remove many good models from the identified set and negatively affect inference. In contrast to their papers, I recommend researchers first use sign restrictions to evaluate the underlying assumptions of the instruments, and only incorporate those that have passed the first stage in the final model.

Finally, my paper contributes to the literature that aims to shrink down the identified set by incorporating more credible information. Tamer (2010) points out that the identified set in a partially-identified model is often too large to be useful for policy analysis. Consequently, a large literature has tried to incorporate information that help sharpen the identified set, including Ludvigson, Ma & Ng (2017), Braun & Brüggemann (2017), Arias, Rubio-Ramírez & Waggoner (2018a), Antolín-Díaz & Rubio-Ramírez (2016), Amir-Ahmadi & Drautzburg (2017), and Amir-Ahmadi & Uhlig (2015). Relative to these papers, I directly model the structural shocks as a linear function of exogenous variables, and hence I can flexibly incorporate many different type of information and exogenous factors. Although my empirical application only uses information from the current value of the instrument, the algorithm can implement sign restrictions on the effect of the instruments on the structural shocks as suggested by Gafarov (2014) or incorporate

[4] Piffer (2017) and Lanne & Luoto (2016) propose Bayesian methods to assess the validity of sign restrictions, but those papers do not discuss instrumental variables.

lags of the instruments to deal with anticipation in rational expectation models as illustrated in Noh (2018).

The rest of this paper is organized as follows. Section 2.2 describes the framework and estimation strategy. Section 2.3 applies the method to investigate the relevance and validity of three monetary policy instruments. Section 2.4 briefly concludes. Additional technical details are shown in the appendices.

2.2 Sign-restricted SVAR with instrumental variables

2.2.1 Model Description

Suppose that the dynamics of the data are summarized by a SVAR(p)

$$\mathbf{A}\mathbf{y}_t = \mathbf{k} + \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{u}_t$$
$$\mathbf{u}_t \sim \text{i.i.d}\, N(\mathbf{0}, \mathbf{D}^*)$$

where $\mathbf{y}_t$ and $\mathbf{u}_t$ are $(n \times 1)$ vector of observed variables and structural shocks at time t. $\mathbf{A}$ is an $(n \times n)$ matrix that governs contemporaneous relationship between observed variables, $\mathbf{B}_i$ $(i = 1, \ldots, p)$ is an $(n \times n)$ matrix of lag coefficients, and $\mathbf{D}^*$ is an $(n \times n)$ diagonal covariance matrix of the structural shocks. For simplicity, the model is rewritten as

$$\mathbf{A}\mathbf{y}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{u}_t \tag{2.1}$$

where $\mathbf{B} \equiv \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \ldots & \mathbf{B}_p & \mathbf{k} \end{bmatrix}$ is $[n \times (np+1)]$ matrix,
and $\mathbf{x}_{t-1} = \begin{bmatrix} \mathbf{y}'_{t-1} & \mathbf{y}'_{t-2} & \cdots & \mathbf{y}'_{t-p} & 1 \end{bmatrix}'$ is $[(np+1) \times 1]$ vector.

Suppose we want to incorporate a $(q \times 1)$ vector of instruments, $\mathbf{z}_t = \left[z_t^{(1)}, z_t^{(2)}, \ldots, z_t^{(q)}\right]$, into the model. We can use the linear projection of the structural shocks on these instruments

$$\mathbf{u}_t = \mathbf{C}\mathbf{z}_t + \mathbf{w}_t \tag{2.2}$$

$$\mathbf{w}_t \sim \text{i.i.d}\, N(\mathbf{0}, \mathbf{D}) \tag{2.3}$$

where $\mathbf{C}$ is a $(n \times q)$ matrix that governs the relationship between the instruments and the endogenous variables, and $\mathbf{D}$ is again assumed to be diagonal. The general model with the instruments can be written as

$$\mathbf{A}\mathbf{y}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{z}_t + \mathbf{w}_t \tag{2.4}$$

Depending on different specifications of $\mathbf{C}$ and the choice of the instruments $\mathbf{z}_t$, model (2.4) can be used to (1) identify one structural shock by one instrument, (2) identify one structural shock by multiple instruments, (3) identify multiple structural shocks by multiple instruments, or (4) incorporate both the current and lagged values of the instruments and other exogenous factors.

The general model (2.4) nests three important special cases regarding the instrument properties. Without loss of generality, suppose we use all instruments to identify the last structural shock in the system. First, if the instruments are *irrelevant but valid*, $\mathbf{C}$ will be restricted to be identically zero. On the other extreme, if we have *relevant but invalid* instruments, $\mathbf{C}$ will be totally unrestricted. Lastly, *relevant and valid* instruments will restrict the elements of the $\mathbf{C}$ matrix to be

$$\mathbf{C} = \begin{bmatrix} \underset{(n-1)\times q}{\mathbf{0}} \\ \underset{1\times q}{\mathbf{c}_n} \end{bmatrix} \tag{2.5}$$

This structure of the matrix $\mathbf{C}$ ensures that the instruments are correlated with only one structural shock (relevant instruments) but not with other shocks (valid instruments).

2.2.2 Discussion

This section briefly discusses the identification problem in SVAR, how researchers use instrumental variables to identify the dynamic causal effects, and how they can use the identifying assumptions from sign restrictions to evaluate the instrument relevance and validity assumptions.

The SVAR(p) described in (2.1) admits a reduced-form VAR(p) representation

$$\mathbf{y}_t = \mathbf{\Pi}_1 \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{2.6}$$

where $\mathbf{\Pi}_1 = \mathbf{A}^{-1}\mathbf{B}$ and $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{\Omega}^*)$ with $\mathbf{\Omega}^* = \mathbf{A}^{-1}\mathbf{D}^* \left(\mathbf{A}^{-1}\right)'$. These parameters are identified from the data and can be consistently estimated by the Ordinary Least Squares (OLS) method. Nevertheless, the parameters $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}^*$ are generally not identified because $\mathbf{A}$ and $\mathbf{D}^*$ may have more parameters than the reduced-form covariance matrix, $\mathbf{\Omega}^*$. Conventional approaches, such as exclusion restrictions, use quantitative restrictions to achieve point-identification. For example, researchers can normalize the elements of the matrix $\mathbf{D}^*$ to one and set the matrix $\mathbf{A}$ to be lower-triangular. Those restrictions impose a particular causal order among the endogenous variables, and hence they are both conceptually and empirically controversial (Ramey (2016), Nakamura & Steinsson (2018)).

Identification by instrumental variables has become a promising alternative recently. To understand the method, consider the reduced-form representation of the SVAR(p) with instrumental variables in (2.4)

$$\mathbf{y}_t = \mathbf{\Pi}_1 \mathbf{x}_{t-1} + \mathbf{\Pi}_2 \mathbf{z}_t + \mathbf{e}_t \tag{2.7}$$

where $\mathbf{\Pi}_1 = \mathbf{A}^{-1}\mathbf{B}$, $\mathbf{\Pi}_2 = \mathbf{A}^{-1}\mathbf{C}$, and $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{\Omega})$ with $\mathbf{\Omega} = \mathbf{A}^{-1}\mathbf{D}\left(\mathbf{A}^{-1}\right)'$. Let $\mathbf{A}^{-1} = \mathbf{\Psi} = \begin{bmatrix} \boldsymbol{\psi}^1 & \boldsymbol{\psi}^2 & \dots & \boldsymbol{\psi}^n \end{bmatrix}$ and consider the specification of $\mathbf{C}$ described in (2.5), we have

$$\mathbf{\Pi}_2 \mathbf{z}_t = \mathbf{\Psi}\mathbf{C}\mathbf{z}_t = \boldsymbol{\psi}^n c_{n,1} z_t^{(1)} + \boldsymbol{\psi}^n c_{n,2} z_t^{(2)} + \dots + \boldsymbol{\psi}^n c_{n,q} z_t^{(q)}$$

If the instruments are relevant and valid, they will identify the last column of $\mathbf{A}^{-1}$, $\boldsymbol{\psi}^n$, up to some scale factors. The last column of $\mathbf{A}^{-1}$ will then be used to infer the dynamic causal effects of the last structural shocks on other macro aggregates. On the other hand, if $\mathbf{C}$ does not satisfy the restrictions in (2.5), the instruments will only identify some non-linear combinations of the structural parameters and hence be useless for dynamic causal inference.

Although instrument validity is crucial to the success of this approach, existing methods have little to say about its legitimacy. When researchers have only one instrument, they cannot determine whether the matrix $\mathbf{C}$, now a vector, satisfies the restrictions in (2.5). Consequently, they often must rely on *ad hoc* arguments to argue that the instrument is ”plausibly exogenous”. And although those restrictions can be tested when researchers use more than one instruments to identify a structural shock, it is not clear what conclusion to draw about the validity of any particular instrument without further information.

Nevertheless, in many cases, researchers do have other identifying assumptions from theory and the empirical literature about reasonable magnitudes and signs of different model quantities. These assumptions can be incorporated as prior belief and sign restrictions on the elements of $\mathbf{A}$ or $\mathbf{A}^{-1}$. Because sign restrictions are grounded in economic theory and robust to model misspecification, this paper proposes to use them as overidentifying assumptions in order to formally test the underlying assumptions of the instruments. To see the intuition of my approach, consider the problem of a researcher who has some doubt about instrument validity. From the Bayesian point of view, her belief after seeing the data can be expressed as an odds ratio between two models: a model where the instrument is valid, and a model where it is not. Conditional on the data, if the first model is more likely than the second, she will conclude that the instrument is valid. Whereas, if the first model is less likely, she will conclude that it is not. Similarly, instrument relevance can be formally assessed by comparing a model where the instrument is assumed to be valid and relevant with a model where the instrument is assumed to be irrelevant.

Formally, let M_0 be a special case of (2.4) where the matrix $\mathbf{C}$ is identically zero, M_1 be the model where the matrix $\mathbf{C}$ satisfies the restrictions in (2.5), and M_2 be the model where the matrix $\mathbf{C}$ is left unrestricted. Thus, M_0 imposes the restrictions that the instruments are not relevant, M_1 imposes the restrictions that the instruments are valid, and M_2 does not impose any restriction. If a Bayesian model comparison favors M_1 over M_0, the instruments are judged to be relevant. Similarly, if a Bayesian model comparison favors M_1 over M_2, the instruments are deemed to be valid.

The robustness of sign restrictions and prior belief play important roles in the success of this method. Wolf (2016) cautions that sign restrictions might identify both the true shocks and mixtures of other shocks that have different dynamics, while Paustian (2007), Canova & Paustian (2011), and Gafarov (2014), and Wolf (2018) assert that the sign-restricted SVAR is more robust and delivers the correct causal inference if researchers are willing to impose an appropriate number of sign restrictions. The required number of sign restrictions would depend on a particular application. In any case, I note that my algorithm can flexibly accommodate sign restrictions in both structural parameters, $\mathbf{A}$ and $\mathbf{C}$, and impact effects of structural shocks, $\mathbf{A}^{-1}$, along with other conventional identifying strategies like short-run and long-run restrictions. Furthermore, in set-identified models such as sign-restricted SVARs, inference is sensitive to prior belief even in large sample. In particular, there exists quantities of which there is no Bayesian updating (Lindley (1957), Poirier (1998)), and Bayesian credible set no longer satisfy frequentist coverage even with infinite amount of data (Moon & Schorfheide (2012)).[5] Following Baumeister & Hamilton (2015), I acknowledge and deal with this problem by using explicit, informative priors carefully constructed from both theory and practice.

[5]Inference in set-identified SVAR is a growing area of research with many notable works, including Gafarov, Meier & Olea (2018), Gafarov, Meier & Olea (2016), Granziera, Moon & Schorfheide (2018), Giacomini, Kitagawa & Volpicella (2017), Giacomini & Kitagawa (2015), Plagborg-Møller (2019), Plagborg-Møller & Wolf (2017), Barnichon, Matthes & Ziegenbein (2016), and Baumeister & Hamilton (2018).

2.2.3 MCMC Algorithm for Estimation

This section describes the priors, the likelihood, and the MCMC algorithm to simulate from the posterior distributions in details. For estimation purposes, the joint prior of $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$ can be decomposed as

$$P(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}) = P(\mathbf{A},\mathbf{C})\,P(\mathbf{D}|\mathbf{A},\mathbf{C})\,P(\mathbf{B}|\mathbf{A},\mathbf{C},\mathbf{D}) \tag{2.8}$$

I allow arbitrary priors on parameters of $\mathbf{A},\mathbf{C}$ and employ natural conjugates for the two conditional priors $P(\mathbf{D}|\mathbf{A},\mathbf{C})$ and $P(\mathbf{B}|\mathbf{A},\mathbf{C},\mathbf{D})$ to ease computation of the posteriors $P(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}|\mathbf{Y}_T)$. Specifically, I use independent inverse-Gamma priors for the variances of the structural shocks

$$p\left(d_{ii}^{-1}|\mathbf{A},\mathbf{C}\right) \sim \Gamma(\kappa_i,\tau_i) \tag{2.9}$$

$$P(\mathbf{D}|\mathbf{A},\mathbf{C}) = \Pi_{i=1}^{n} p\left(d_{ii}^{-1}|\mathbf{A},\mathbf{C}\right) \tag{2.10}$$

where $\Gamma(\kappa_i,\tau_i)$ denotes the Gamma distribution with parameters κ_i and τ_i. For the lag parameters, I use multivariate normal priors that are independent across equations

$$P(\mathbf{b}_i|\mathbf{A},\mathbf{C},\mathbf{D}) \sim N(\mathbf{m}_i, d_{ii}\mathbf{M}_i) \tag{2.11}$$

$$P(\mathbf{B}|\mathbf{A},\mathbf{C},\mathbf{D}) = \Pi_{i=1}^{n} P(\mathbf{b}_i|\mathbf{A},\mathbf{C},\mathbf{D}) \tag{2.12}$$

where $N(\mathbf{m}_i, d_{ii}\mathbf{M}_i)$ denotes the multivariate normal distribution with mean $\mathbf{m}_i$ and covariance matrix $d_{ii}\mathbf{M}_i$. The overall prior is

$$P(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}) = P(\mathbf{A},\mathbf{C})\,\Pi_{i=1}^{n} p\left(d_{ii}^{-1}|\mathbf{A},\mathbf{C}\right)\Pi_{i=1}^{n} P(\mathbf{b}_i|\mathbf{A},\mathbf{C},\mathbf{D}) \tag{2.13}$$

I condition on the instrumental variables in the estimation and make use of the conditional likelihood, which is [6]

$$P(\mathbf{Y}_T|\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{Z}_T) = (2\pi)^{-Tn/2}\det|\mathbf{A}|^T \det|\mathbf{D}|^{-T/2} \times \exp\left[-(1/2)\sum_{t=1}^{T}(\mathbf{A}\mathbf{y}_t - \mathbf{C}\mathbf{z}_t - \mathbf{B}\mathbf{x}_{t-1})'\mathbf{D}^{-1}(\mathbf{A}\mathbf{y}_t - \mathbf{C}\mathbf{z}_t - \mathbf{B}\mathbf{x}_{t-1})\right] \tag{2.14}$$

Given the priors and the likelihood, the posteriors will be characterized by Bayes' rule. The joint posterior of **A** and **C** is calculated by the random-walk Metropolis-Hasting algorithm, while the posteriors of **B** and **D** are their respective natural conjugates. The estimation procedure is stated formally below.

Proposition 1. *Let the priors be given as in* (2.8)-(2.13) *and the likelihood function be given as in* (2.14). *Moreover, let* $\mathbf{a}_i'$ *denote the i-th row of* $\mathbf{A}$*,* $\mathbf{c}_i'$ *denote the i-th row of* $\mathbf{C}$*,* $\mathbf{P}_i$ *denote the Cholesky factor of* $\mathbf{M}_i^{-1} = \mathbf{P}_i\mathbf{P}_i'$*. Then, the posteriors are*

$$P(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}|\mathbf{Y}_T,\mathbf{Z}_T) = P(\mathbf{A},\mathbf{C}|\mathbf{Y}_T,\mathbf{Z}_T)\,\Pi_{i=1}^{n} p\left(d_{ii}^{-1}|\mathbf{A},\mathbf{C},\mathbf{Y}_T,\mathbf{Z}_T\right)\Pi_{i=1}^{n} P(\mathbf{b}_i|\mathbf{A},\mathbf{C},\mathbf{D},\mathbf{Y}_T,\mathbf{Z}_T) \tag{2.15}$$

[6]The motivation for using conditional likelihood is provided in Appendix B.1.

with

$$P(\mathbf{A},\mathbf{C}|\mathbf{Y}_T,\mathbf{Z}_T) = \frac{k_T P(\mathbf{A},\mathbf{C})\left[\det\left(\mathbf{A}\hat{\boldsymbol{\Omega}}\mathbf{A}'\right)\right]^{T/2}}{\Pi_{i=1}^{n}\left[2\tau_i^*/T\right]^{\kappa_i^*}} \tag{2.16}$$

$$p\left(d_{ii}^{-1}|\mathbf{A},\mathbf{C},\mathbf{Y}_T,\mathbf{Z}_T\right) \sim \Gamma(\kappa_i^*,\tau_i^*) \tag{2.17}$$

$$P(\mathbf{b}_i|\mathbf{A},\mathbf{C},\mathbf{D},\mathbf{Y}_T,\mathbf{Z}_T) \sim N(\mathbf{m}_i^*, d_{ii}\mathbf{M}_i^*) \tag{2.18}$$

$$\hat{\boldsymbol{\Omega}} = T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t' \tag{2.19}$$

$$\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\boldsymbol{\Phi}}\mathbf{x}_{t-1} \tag{2.20}$$

$$\hat{\boldsymbol{\Phi}} = \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_{t-1}'\right)\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_{t-1}'\right)' \tag{2.21}$$

$$\kappa_i^* = \kappa_i + (T/2) \tag{2.22}$$

$$\tau_i^* = \tau_i + (\zeta_i^*/2) \tag{2.23}$$

$$\zeta_i^* = \left(\tilde{\mathbf{Y}}_i'\tilde{\mathbf{Y}}_i\right) - \left(\tilde{\mathbf{Y}}_i'\tilde{\mathbf{X}}_i\right)\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i\right) \tag{2.24}$$

$$\mathbf{m}_i^* = \left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i\right) \tag{2.25}$$

$$\mathbf{M}_i^* = \left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \tag{2.26}$$

$$\tilde{\mathbf{Y}}_i = \begin{bmatrix}\mathbf{y}_1'\mathbf{a}_i - \mathbf{c}_i'\mathbf{z}_{i1} & \ldots & \mathbf{y}_T'\mathbf{a}_i - \mathbf{c}_i'\mathbf{z}_{iT} & \mathbf{m}_i'\mathbf{P}_i\end{bmatrix}' \tag{2.27}$$

$$\tilde{\mathbf{X}}_i = \begin{bmatrix}\mathbf{x}_0' & \ldots & \mathbf{x}_{T-1}' & \mathbf{P}_i\end{bmatrix}' \tag{2.28}$$

The differences between this proposition and that in Baumeister & Hamilton (2015) are equations (2.16) and (2.27). In particular, the target for the Metropolis-Hasting algorithm in (2.16) takes into account the priors for the matrix $\mathbf{C}$, and (2.27) redefines $\tilde{\mathbf{Y}}_i$ to account for the instrumental variables $\mathbf{z}_i$. The estimation procedure and its proof are identical to those in their paper.

2.2.4 Bayesian Model Comparison

Bayes factor

Bayesian model comparison is used to evaluate instrument relevance and validity. Generally, a Bayesian model comparison between model k and k' is done by calculating the posterior odds ratio, showing which model is more likely now that we have seen the data.[7] This can be written as the product of the Bayes factor and the prior odds ratio.

$$\frac{P(M_k|\mathbf{Y}_T,\mathbf{Z}_T)}{P(M_{k'}|\mathbf{Y}_T,\mathbf{Z}_T)} = \frac{P(\mathbf{Y}_T|\mathbf{Z}_T,M_k)}{P(\mathbf{Y}_T|,\mathbf{Z}_T,M_{k'})}\frac{P(M_k)}{P(M_{k'})} \tag{2.29}$$

Suppose before seeing the data, we assume model k is as likely as model k', then the prior odds ratio (i.e. $P(M_k)/P(M_{k'})$) will be one and equation (2.29) simplifies to

$$\frac{P(M_k|\mathbf{Y}_T,\mathbf{Z}_T)}{P(M_{k'}|\mathbf{Y}_T,\mathbf{Z}_T)} = \frac{P(\mathbf{Y}_T|\mathbf{Z}_T,M_k)}{P(\mathbf{Y}_T|,\mathbf{Z}_T,M_{k'})} \tag{2.30}$$

The quantity on the right-hand side of equation (2.30) is the so-called Bayes factor, which is the ratio of two marginal likelihoods. The Bayes factor between model k and model k' is denoted as

$$B_{k,k'} = \frac{P(M_k|\mathbf{Y}_T,\mathbf{Z}_T)}{P(M_{k'}|\mathbf{Y}_T,\mathbf{Z}_T)} = \frac{P(\mathbf{Y}_T|\mathbf{Z}_T,M_k)}{P(\mathbf{Y}_T|,\mathbf{Z}_T,M_{k'})}$$

The Bayes factor shows how likely model k is relative to model k'. For example, the value of two means that model k is twice as likely as model k' after the data are observed.

[7] Since I condition on the instruments, they don't affect the calculation of the odds ratio. Appendix B.1 provides more details.

Marginal likelihood estimation

The Bayes factor is the ratio of two high dimensional integrals. Let $\boldsymbol{\theta}$ denote all unknown parameters. The marginal likelihood of model k is defined as

$$P(\mathbf{Y}_T|\mathbf{Z}_T, M_k) = \int P(\mathbf{Y}_T|\boldsymbol{\theta}, \mathbf{Z}_T, M_k) P(\boldsymbol{\theta}|M_k)\, \mathrm{d}\boldsymbol{\theta} \tag{2.31}$$

where $P(\mathbf{Y}_T|\boldsymbol{\theta}, \mathbf{Z}_T, M_k)$ is the likelihood of model M_k, and $P(\boldsymbol{\theta}|M_k)$ is the prior, which is assumed to be independent of $\mathbf{Z}_T$.[8] Let $f(\boldsymbol{\theta})$ be a known multivariate density function of $\boldsymbol{\theta}$. Bayes' theorem implies that

$$\frac{1}{P(\mathbf{Y}_T|\mathbf{Z}_T, M_k)} = \int \frac{f(\boldsymbol{\theta})}{P(\mathbf{Y}_T|\boldsymbol{\theta}, \mathbf{Z}_T, M_k) P(\boldsymbol{\theta}|M_k)} P(\boldsymbol{\theta}|\mathbf{Y}_T, \mathbf{Z}_T, M_k)\, \mathrm{d}\boldsymbol{\theta}$$

Thus, a natural estimator for the marginal likelihood from the posterior draws would be

$$\hat{P}(\mathbf{Y}_T|\mathbf{Z}_T, M_k) = \left[\frac{1}{N_0}\sum_{n_0=1}^{N_0} \frac{f(\boldsymbol{\theta}^{n_0})}{P(\mathbf{Y}_T|\boldsymbol{\theta}^{n_0}, \mathbf{Z}_T, M_k) P(\boldsymbol{\theta}^{n_0}|M_k)}\right]^{-1} \tag{2.32}$$

where N_0 is the number of posterior draws after discarding the burn-in sample. The choice of $f(\boldsymbol{\theta})$ is important in the calculation of the marginal likelihood. For instance, if we choose $f(\boldsymbol{\theta})$ to be the prior distribution (i.e. $f(\boldsymbol{\theta}) = P(\boldsymbol{\theta}|M_k)$), we will have a harmonic mean estimator

$$\hat{P}_{HMM}(\mathbf{Y}_T|\mathbf{Z}_T, M_k) = \left[\frac{1}{N_0}\sum_{n_0=1}^{N_0} \frac{1}{P(\mathbf{Y}_T|\boldsymbol{\theta}^{n_0}, \mathbf{Z}_T, M_k)}\right]^{-1}$$

However, this estimator has infinite variance and is numerically inefficient. For the method to work well, $f(\boldsymbol{\theta})$ needs to be a good approximation of the posterior distribution and has a thinner tail than the posterior kernel, $P(\mathbf{Y}_T|\boldsymbol{\theta}, \mathbf{Z}_T, M_k) P(\boldsymbol{\theta}|M_k)$, to ensure convergence of the Monte Carlo average. Geweke (1999) proposes to use a truncated normal distribution, while Sims, Waggoner & Zha (2008) constructs a more sophisticated choice for $f(\boldsymbol{\theta})$. Appendix B.2 describes

[8] $\boldsymbol{\theta} = (vec(\mathbf{A}), vec(\mathbf{B}), vec(\mathbf{C}), vec(\mathrm{diag}(\mathbf{D})))'$

those two choices in details.

2.3 Application: Three instruments for monetary shock

This section applies the above method to investigate the relevance and validity of three instruments for monetary shocks. It also shows how my method can combine information from a relevant and valid instrument with sign restrictions to improve dynamic causal inference in SVAR. I first describe the data then present the analysis from both frequentist and Bayesian perspectives to illustrate the benefits of my approach.

2.3.1 Data Description

The data are publicly available and are described in detail in Appendix B.3. Macroeconomic time series are downloaded from FRED, while monetary policy instruments are collected from Mark Watson and Yuriy Gorodnichenko's website.[9] The sample period is 1954Q3 to 2008Q4. Four *quarterly* time series are used: real GDP, real potential GDP, personal consumption expenditures deflator (PCE deflator), and effective federal funds rate. The output gap is calculated as the log difference between real and potential GDP, and the inflation rate is computed as the Y/Y change of the PCE deflator. The instruments are similar to those in Stock & Watson (2012a); they come from Romer & Romer (2004)'s narrative method, Smets & Wouters (2007)'s DSGE model, and Sims & Zha (2006)'s Markov-Switching SVAR. Missing values of the instruments are replaced with zeros as done in Romer & Romer (2004).

Figure 2.1 plots the macroeconomic aggregates and the instruments. It shows that both macroeconomic variables and monetary policy instruments are less volatile after the 1980s, which is consistent with Stock & Watson (2002)'s documentation of the Great Moderation. A dampening in the volatility of monetary policy instruments is also consistent with Ramey (2016)'s observation

[9]As a robustness check, I also consider two other updated Romer-Romer monetary policy instruments from Wieland & Yang (2015) The results are qualitatively the same.

that monetary policy has been conducted in a more systematic manner after 1980, and thus true monetary policy surprises are hard to identify from the data.

Table 2.1 shows the summary statistics of the data, and Table 2.2 displays the pairwise cross-correlation of the instruments. Table 2.2 paints a similar picture to that in Stock & Watson (2012a), namely monetary policy instruments are not all highly correlated with each other. The Romer-Romer instrument appears to be highly correlated with Sims-Zha instrument and less correlated with that of Smets-Wouters. Although the correlations between instruments are not the same as the correlations between their predictive shocks, their lack of correlation is suggestive evidence that they might not identify the same shock. One possible explanation is that different instruments capture different dimensions of monetary policy shocks, and another explanation is that some of the instruments are contaminated by other contemporaneous shocks and hence invalid. In the following investigation, I will assume that all three monetary policy instruments identify the same monetary policy shock and use a sign-restricted SVAR model to study the latter explanation.

Table 2.1: **Summary statistics**. This table shows the summary statistics of the variables in the empirical application. Descriptions of the data and their availability are explained in the text. There are 218 quarterly observations for each variable in the period between 1954Q3 and 2008Q4. The units are all in percentage points.

Summary statistics				
Variables	**Mean**	**Standard Deviations**	**Minimum**	**Maximum**
Output gap	-0.5	2.2	-7.6	5.5
Inflation rates	3.4	2.3	-0.2	10.9
Fed Funds rates	5.6	3.3	0.5	17.8
Romer-Romer monetary instrument	0.0	0.5	-4.1	2.5
Sims-Zha monetary instrument	0.0	2.4	-15.3	14.9
Smets-Wouters monetary instrument	-0.2	0.9	-3.6	4.8

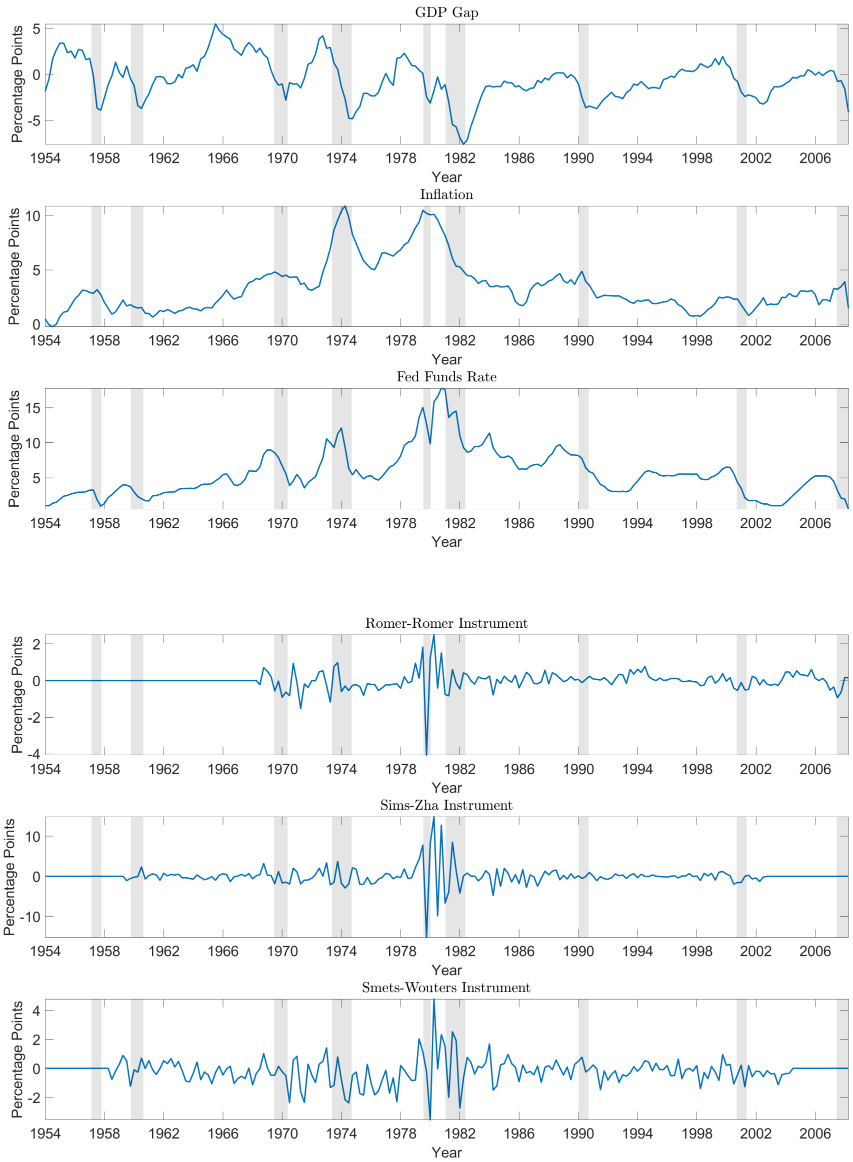

Figure 2.1: **Output gap, inflation, fed funds rates, together with Romer-Romer, Sims-Zha, and Smets-Wouters instruments**. Output gap is the log difference between real and potential GDP, multiplied by 100. Inflation rate is the log difference between the Y/Y change of the PCE deflator, multiplied by 100. Fed funds rates is the Effective Federal Funds Rate. Romer and Romer's instruments are the residuals from the regression between shocks constructed by the narrative methods on the Fed's Greenbook forecasts of output and inflation. Sims and Zha's instruments are shocks constructed from the VAR that includes Markov-Switching variances and no time-varying parameters. Smets and Wouters' instruments are interest rate shocks as calculated from Smets and Wouters' DSGE model. Shaded area indicates NBER recession periods. More detailed descriptions and data sources are in Appendix B.3.

Table 2.2: **Cross-correlations**: This table shows the cross-correlation of the instruments for monetary policy shocks. Descriptions of the data and their availability are explained in the text.

Cross-correlations of the instruments (1954Q3-2008Q4)			
Variables	Romer-Romer instrument	Sims-Zha instrument	Smets-Wouters instrument
Romer-Romer instrument	1.00		
Sims-Zha instrument	0.76	1.00	
Smets-Wouters instrument	0.37	0.44	1.00

2.3.2 Proxy SVAR and tests of overidentifying restrictions

This section uses frequentist methods to analyze instrument relevance and validity. To implement the Proxy SVAR, Stock & Watson (2018) recommend the following IV regression

$$y_t = \alpha_0 + \alpha_1 r_t + \mathbf{b}'\mathbf{x}_{t-1} + \varepsilon_t$$
$$r_t = \gamma_0 + \mathbf{c}'\mathbf{z}_t + \mathbf{d}'\mathbf{x}_{t-1} + \zeta_t$$

where y_t is the output gap, r_t is the effective fed funds rate, $\mathbf{z}_t = \left(z_{1t}, z_{2t}, \ldots, z_{qt}\right)'$ is a $(q \times 1)$ vector of monetary shocks instruments, and $\mathbf{x}_{t-1} = \left(\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-p}\right)'$is a $(k \times 1)$ vector of lag variables in the VAR. Let $\mathbf{u}_t = \left(u^s_t, u^d_t, u^m_t\right)$ denote the supply shock, the demand shock, and the monetary policy shocks respectively. Stock & Watson (2018) show that we can consistently estimate α_1 and use it to study the dynamic causal effect of monetary policy shock on the output gap under three assumptions

1. Instrument relevance: $\mathbb{E}\left(u^m_t \mathbf{z}_t\right) \neq 0$
2. Instrument validity: $\mathbb{E}\left(u^s_t \mathbf{z}_t\right) = 0$ and $\mathbb{E}\left(u^d_t \mathbf{z}_t\right) = 0$
3. Invertibility: $\mathbf{v}_t = \mathbf{Q}\mathbf{u}_t$ where $\mathbf{v}_t$ is the innovations in the reduced-form VAR and $\mathbf{Q}$ is a (3×3) rotation matrix

Assumption (1) and (2) are standard assumptions in IV regressions to achieve consistency of the estimator, while assumption (3) is required in the SVAR setting to ensure that α_1 can be used to infer the dynamic effect of monetary policy shock. Intuitively, assumption (3) means that monetary policy shocks can be recovered from the current and past information of the output gap, inflation, and fed funds rate. Although Stock & Watson (2018) caution that invertibility might fail when the variables in the VAR are not sufficient to account for the Fed information set or when economic agents are forward-looking, Wolf (2018) shows that monetary policy shocks are nearly invertible even for a standard three-variable VAR. In this application, I will only focus on (1) and (2).

To analyze instrument relevance, I first estimate the above IV regression and let each instrument enter one by one. Table 2.3 reports the partial R-squared, F-statistics, heteroskedasticity-robust F-statistics, and their corresponding p-values for all three instruments.[10] Both the normal and robust F-statistics are significantly higher than 10, and their p-values are indistinguishable from zero. Thus, the empirical evidence suggests that all three instruments are strong. I note that the first-stage F-statistics are higher than those found in Stock & Watson (2012a). One potential explanation is that their study employs a Factor-SVAR model with more than 200 macroeconomic time series whereas my model only has three time series.

Table 2.3: **First-stage statistics in SVAR-IV model**. This table shows the partial R-squared, F-statistics, Robust F-statistics, and their corresponding p-values in the first-stage regression of the SVAR-IV model. The p-values for both F-statistics and Robust F-statistics are both close to 0, thus I only show one column for both.

First-stage F-statistics for three different instruments (1954Q3-2008Q4)				
Variables	**Partial R-squared**	**F-statistics**	**Robust F-statistics**	**p-values**
Romer-Romer shocks	0.37	119	19	0.00
Sims-Zha shocks	0.57	262	34	0.00
Smets-Wouters shocks	0.50	203	52	0.00

Next, I use tests of overidentifying restrictions to investigate the validity of the instruments. I conduct two separate exercises. The first uses all three instruments at once, and the second

[10]The partial R-squared measures how much the variation in the dependent variable is explained by the variation of the instruments excluded other exogenous variables.

uses two instruments at a time. I use three different tests, which are developed by Sargan (1958), Basmann (1960), and Wooldridge (1995). The first two tests assume that the error-terms are i.i.d., whereas the last one controls for heteroskedastic errors. The null hypothesis is that all three instruments are valid. Table 2.4 reports the test statistics and their corresponding p-values for the first exercise. The Sargan's score and the Basmann's test deliver p-values that are close to zero, while the Wooldrige's score has a p-value around 6%. Hence, the overall results suggest that at least one instrument is invalid.

Table 2.4: **Test of overidentifying restrictions in SVAR-IV model-all three instruments**. This table shows the test statistics from three different tests of overidentifying restrictions: Sargan (1958), Basmann (1960), and Wooldridge (1995), together with their corresponding p-values.

Test of overidentifying restrictions-all three instruments		
Test	**Statistics**	**p-values**
Sargan's score	10.81	0.0045
Basmann's test	10.53	0.0052
Wooldrige's score	5.62	0.0601

Table 2.5 summarizes the results from the second exercise and sheds new light on the first results. All three tests fail to reject the null hypothesis when the Romer-Romer and Sims-Zha instruments are used together, however, they strongly reject for the other two cases. A failure to reject in overidentifying-restriction tests does not imply that the instruments are valid but only implies that the IV estimates from those instruments are similar to each other. Thus, the results suggest that the Romer-Romer and Sims-Zha instruments deliver similar IV estimates, which are different from those estimated by the Smets-Wouters instrument.

In summary, frequentist evidence reveals that all three instruments are strong but at least one of them is not valid. However, they cannot say which instrument is valid and which one is not because the null hypothesis that only one instrument is valid is untestable under the frequentist paradigm.

Table 2.5: **Test of overidentifying restrictions in SVAR-IV model-two instruments**. This table shows the test statistics from three different tests of overidentifying restrictions: Sargan (1958), Basmann (1960), and Wooldridge (1995), together with their corresponding p-values.

Test of overidentifying restrictions-two instruments						
	Romer-Romer/Sims-Zha		**Romer-Romer/Smets-Wouters**		**Sims-Zha/Smets-Wouters**	
Test	**Statistics**	**p-values**	**Statistics**	**p-values**	**Statistics**	**p-values**
Sargan's score	0.11	0.74	7.17	0.01	13.05	0.00
Basmann's test	0.11	0.75	6.90	0.01	12.92	0.00
Wooldrige's score	0.14	0.71	4.10	0.04	9.27	0.00

2.3.3 A Bayesian proxy SVAR of monetary policy

This section applies my Bayesian approach to investigate instrument relevance and validity. I first introduce the trivariate system of monetary policy in Baumeister & Hamilton (2018) which consists of the output gap, inflation, and nominal fed funds rate. Let y_t be the output gap, π_t be the inflation rate, r_t be the effective fed funds rate, z_t be the monetary policy instrument (i.e. Romer-Romer instrument, Sims-Zha instrument, or Smets-Wouters instrument), and $\mathbf{x}_{t-1}$ be the vector of lag variables in the SVAR, the model is characterized by the following structural equations

1) The Phillips curve:

$$y_t = k^s + \alpha^s \pi_t + [\mathbf{b}^s]' \mathbf{x}_{t-1} + u_t^s \tag{2.33}$$

2) The aggregate demand equation:

$$y_t = k^d + \beta^d \pi_t + \gamma^d r_t + \left[\mathbf{b}^d\right]' \mathbf{x}_{t-1} + u_t^d \tag{2.34}$$

3) The Monetary Policy Rule:

$$r_t = k^m + (1-\rho)\,\psi^y y_t + (1-\rho)\,\psi^\pi \pi_t + +\rho r_{t-1} + [\mathbf{b}^m]' \mathbf{x}_{t-1} + u_t^m \tag{2.35}$$

I impose four sign restrictions on the structural parameters: (1) the Phillips curve is downward sloping $(\alpha^s > 0)$, (2) raising interest rate will not stimulate aggregate demand $\left(\gamma^d < 0\right)$, (3) the Fed will raise interest rate when the inflation rate is higher than its target$(\psi^y > 0)$ or the output gap is higher than its potential $(\psi^\pi > 0)$, and (4) the Fed wants to increase its interest rate smoothly over time $(0 < \rho < 1)$. I do not use any prior on the impact effect of monetary policy shocks on those three variables because these effects should be identified by the instruments. Together, equation (2.33), (2.34), and (2.35) constitute the baseline model M_0 where the instrument is irrelevant.

If one has an instrument for monetary policy shock, one can model the monetary policy shock as

4) Monetary Policy shock:

$$u_t^m = \chi^m z_t + w_t^m \tag{2.36}$$

where χ^m measures the effect of the instrument on the monetary policy shock, and w^m is the part of the monetary policy shock that is orthogonal to the instrument as well as other shocks in the system. Equation (2.33), (2.34), (2.35), and (2.36) constitute the first alternative model M_1 where the instrument is assumed to be valid.

Furthermore, if one has doubt about the validity of the instrument, one can investigate that assumption by augmenting model M_1 with two additional equations

5) Supply shock:

$$u_t^s = \chi^s z_t + w_t^s \tag{2.37}$$

6) Demand shock:

$$u_t^d = \chi^d z_t + w_t^d \tag{2.38}$$

where χ^s and χ^d measure the effects of the instruments on supply and demand shocks respectively. If either χ^s or χ^d is different from zero, that will be evidence against instrument validity. The shocks w^m, w^s, and w^d are assumed to be orthogonal to each other as in the general formulation. Thus, the model where the instrument is allowed to be invalid, M_2, is summarized by six equations

(2.33), (2.34), (2.35), (2.36), (2.37), and (2.38).

Priors for contemporaneous coefficients and shocks

This section discusses the priors for the contemporaneous parameters and the coefficients in the monetary, supply, and demand shock equations. I assume that the structural parameters in $\mathbf{A},\mathbf{C}$ are independent of each other. In particular, the priors are

1. Baseline model where the instruments are restricted to be irrelevant (M_0):

$$P(\mathbf{A},\mathbf{C}) = p(\alpha^s)\, p\left(\beta^d\right) p\left(\gamma^d\right) p(\psi^y)\, p(\psi^\pi)\, p(\rho)$$

2. Model where the instrument is assumed to be valid and generally relevant (M_1):

$$P(\mathbf{A},\mathbf{C}) = p(\alpha^s)\, p\left(\beta^d\right) p\left(\gamma^d\right) p(\psi^y)\, p(\psi^\pi)\, p(\rho)\, p(\chi^m)$$

3. Model where the instrument is allowed to be invalid (M_2):

$$P(\mathbf{A},\mathbf{C}) = p(\alpha^s)\, p\left(\beta^d\right) p\left(\gamma^d\right) p(\psi^y)\, p(\psi^\pi)\, p(\rho)\, p(\chi^s)\, p\left(\chi^d\right) p(\chi^m)$$

I follow Baumeister & Hamilton (2018) in using additional prior information about elements of $\mathbf{A}$ and do not repeat their motivations in this paper. My main innovation here is the specification of the matrix $\mathbf{C}$. I set the priors for all elements of $\mathbf{C}$ to be centered at zero with a small variance. I use symmetric priors because it is hard to tell how the instruments will affect the structural shocks before seeing the data. Also, because the main interest is whether those parameters are different from zero, a tight prior centered at zero will help avoid the Bartlett-Jeffreys-Lindley's paradox (Jeffreys (1967), Lindley (1957), Bartlett (1957)). The phenomenon occurs when a sharp null hypothesis is rejected by frequentist methods but nonetheless received a high posterior odds based on a Bayesian analysis with small prior probability for the null and diffuse prior for the alternative. To avoid this apparent disagreement between frequentist and Bayesian approaches, Kass & Raftery (1995) recommend using a proper prior with small scale for the tested parameter, essentially ruling out the use of improper priors. In this application, I set the prior belief for χ^m to be a student t distribution centered at 0 with scale 0.2 and 3 degrees of freedom in model M_1. This prior is proper, has a small variance, and allows for the possibility of a weak instrument. Similarly, I use that same prior for all three parameters χ^s, χ^d, and χ^m in model M_2. Thus, these proper priors allow for the possibility of an invalid instrument.

Priors for the covariance matrix

This section describes the prior for $\mathbf{B},\mathbf{D}$ conditional on $\mathbf{A},\mathbf{C}$. They are similar for all three models. As in the general formulation, I set the prior belief for elements of the covariance matrix to be independent from each other

$$P(\mathbf{D}|\mathbf{A},\mathbf{C}) = \prod_{i=1}^{3} p\left(d_{ii}^{-1}|\mathbf{A},\mathbf{C}\right)$$

where each of the element d_{ii} follows an inverse-Gamma distribution

$$p\left(d_{ii}^{-1}|\mathbf{A},\mathbf{C}\right) = \begin{cases} \dfrac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)} & \left(d_{ii}^{-1}\right)^{k_i-1} \exp\left(-\tau_i d_{ii}^{-1}\right) \quad \text{for } d_{ii}^{-1} \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

In this specification, $\dfrac{\kappa_i}{\tau_i}$ is the prior mean for d_{ii}^{-1} , and $\dfrac{\kappa_i}{\tau_i^2}$ is the prior variance. Since the prior of the variances should reflect the scale of the data, I first fit three univariate $AR(4)$ model to the data

$y_t = \beta_{10} + \sum_{i=1}^{4} \beta_{1i} y_{t-i} + e_{1t}$

$\pi_t = \beta_{20} + \sum_{i=1}^{4} \beta_{2i} \pi_{t-i} + e_{2t}$

$r_t = \beta_{30} + \sum_{i=1}^{4} \beta_{3i} r_{t-i} + e_{3t}$

Then, I calculate the fitted residuals from those regressions: $\hat{e}_t = \begin{bmatrix} \hat{e}_{1t} & \hat{e}_{2t} & \hat{e}_{3t} \end{bmatrix}'$, and estimate the sample covariance matrix:

$\hat{\mathbf{S}} = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t \hat{e}_t'$

Finally, I set $\kappa_i = 2$ for all i, which gives my prior a weight that equals to four observations of the data in the posterior. Next, I set the prior mean for d_{ii}^{-1} to be the reciprocal of the i^{th} diagonal element of $\mathbf{A}\hat{\mathbf{S}}\mathbf{A}$. In sum, the prior density for elements of the covariance matrix will be

$$p\left(d_{ii}^{-1} | \mathbf{A}, \mathbf{C}\right) \sim \Gamma\left(2, 2\mathbf{a}_i' \hat{\mathbf{S}} \mathbf{a}_i\right)$$

Priors for the lag coefficients

I set the priors for the structural lag coefficients in such a way that the reduced-form lag coefficients are consistent with a Minnesota prior. Specifically, the priors for lag coefficients are assumed to be independent across equations

$$P(\mathbf{B}|\mathbf{A}, \mathbf{C}, \mathbf{D}) = \prod_{i=1}^{3} P(\mathbf{b}_i | \mathbf{A}, \mathbf{C}, \mathbf{D})$$

where $P(\mathbf{b}_i | \mathbf{A}, \mathbf{C}, \mathbf{D}) \sim N(\mathbf{m}_i, d_{ii} \mathbf{M}_i)$. For the mean, I set $\mathbf{m}_i(\alpha) = \eta' \mathbf{a}_i$ where $\eta = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times 10} \end{bmatrix}$.[11] And for the covariance matrix $\mathbf{M}_i$, let $\sqrt{s_{ii}}$ be the estimated standard deviation of the $AR(4)$ that

[11] Note that if the reduced-form lag coefficients (Φ) follow the Minnesota prior (meaning $\mathbb{E}(\boldsymbol{\Phi}) = \eta$). Then, $\mathbb{E}(\mathbf{B}|\mathbf{A}) = \mathbb{E}(\mathbf{A}\Phi|\mathbf{A}) = \mathbf{A}\mathbb{E}(\Phi|\mathbf{A}) = \eta\mathbf{A}$. Thus, if the prior for the reduced-form lag coefficients are the Minnesota prior, then the prior for the lag coefficients in the structural model are normal with mean $\mathbf{m}_i(\alpha) = \eta' \mathbf{a}_i$.

fits to variable i. I define:

$$\mathbf{v}_1' = \begin{bmatrix} \frac{1}{1^{2\lambda_1}} & \frac{1}{2^{2\lambda_1}} & \frac{1}{3^{2\lambda_1}} & \frac{1}{4^{2\lambda_1}} \end{bmatrix}$$
$$\mathbf{v}_2' = \begin{bmatrix} s_{11}^{-1} & s_{22}^{-1} & s_{33}^{-1} \end{bmatrix}$$
$$\mathbf{v_3} = \lambda_0^2 \begin{bmatrix} \mathbf{v}_1 \mathbf{v}_2 \\ \lambda_3^2 \end{bmatrix}$$

Then, $\mathbf{M}_i$ will be the diagonal matrix whose row r column r element is the r^{th} element of $\mathbf{v}_3$: $M_{i,rr} = v_{3r}$. Intuitively, we are letting coefficients on higher lags to shrink to zero by setting decreasing values for diagonal elements of $\mathbf{M}_i$. The hyper-parameter λ_0 captures how much confidence we have in the prior. A higher value implies a higher variance and less confidence. The hyper-parameter λ_1 governs how quickly the coefficients shrink to zero. And the hyper-parameter λ_3 describes the confidence in the prior for the constant, the higher the value, the less confidence we have. I set $\lambda_0 = 0.2$, $\lambda_1 = 1$, and $\lambda_3 = 100$.

In addition, the third element of the monetary policy rule equation is expected to be close to ρ. This prior information is captured using a prior for the third element of $\mathbf{b}_3$ as follows

$$\rho = \mathbf{I}_{13}^{(3)} \mathbf{b}_3 + v_3 \tag{2.39}$$

where $v_3 \sim N(0, d_{33}V_3)$ and $\mathbf{I}_{13}^{(3)}$ is the third row of $\mathbf{I}_{13}$. The variance parameter, V_3, reflects the strength of the belief in the sense that the smaller V_3 is, the more likely that parameter is close to ρ. In the application, I set $V_3 = 0.1$. To estimate the model with this information, the new $\tilde{\mathbf{Y}}_3$ and

$\tilde{\mathbf{X}}_3$ are modified as

$$\tilde{\mathbf{Y}}_3 = \begin{bmatrix} \mathbf{a}_3'\mathbf{y}_1 - \chi^m z_1 & \dots & \mathbf{a}_3'\mathbf{y}_T - \chi^m z_T & \mathbf{m}_3'\mathbf{P}_3 & \frac{\rho}{\sqrt{V_3}} \end{bmatrix}'$$
$$\tilde{\mathbf{X}}_3 = \begin{bmatrix} \mathbf{x}_0 & \dots & \mathbf{x}_{T-1} & \mathbf{P}_3 & \frac{\mathbf{I}_{13}^{(3)}}{\sqrt{V_3}} \end{bmatrix}'$$

Tables 2.6, 2.7, 2.8 provide complete summaries for the priors of model M_0, M_1, and M_2 respectively.

Table 2.6: **Prior distributions of the baseline model** (M_0). This table shows the prior distribution of **A**, **D**, and **B** together with their hyper-parameters. For Student t distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

Parameter	Meaning	Location	Scale	Skew	Sign restriction
	Priors for contemporaneous coefficients of A				
	Student t distribution with 3 degrees of freedom				
α^s	Effect of π on supply	2	0.4	–	$\alpha^s \geq 0$
β^d	Effect of π on demand	0.75	0.4	–	None
γ^d	Effect of r on demand	-1	0.4	–	$\gamma^d \leq 0$
ψ^y	Fed response to y	0.5	0.4	–	$\psi^y \geq 0$
ψ^π	Fed response to π	1.5	0.4	–	$\psi^\pi \geq 0$
	Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$				
ρ	Interest rate smoothing	0.5	0.2	–	$0 \leq \rho \leq 1$
	Priors for structural variances $\mathbf{D}\|\mathbf{A}$				
	Gamma distribution				
d_{ii}^{-1}	Reciprocal of variance	$1/\left(\mathbf{a}_i'\hat{\mathbf{S}}\mathbf{a}_i\right)$	$1/\left(\sqrt{2}\mathbf{a}_i'\hat{\mathbf{S}}\mathbf{a}_i\right)$	–	$d_{ii} > 0$
	Priors for lag coefficients $\mathbf{B}\|\mathbf{A},\mathbf{D}$				
	Normal distribution				
$\mathbf{b}_i$	Lagged coefficients of equation i	$\eta'\mathbf{a}_i$	$\sqrt{d_{ii}\mathbf{M}_i}$	–	None
In addition,					
b_{33}	Third element of monetary equation	ρ	$\sqrt{d_{33}/10}$	–	None

Table 2.7: **Prior distributions of the model with a relevant and valid instrument** (M_1). This table shows the prior distribution of **A**, **D**, **B**, and **C** together with their hyper-parameters. For Student t distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

Parameter	Meaning	Location	Scale	Skew	Sign restriction
	Priors for contemporaneous coefficients of A				
	Student t distribution with 3 degrees of freedom				
α^s	Effect of π on supply	2	0.4	–	$\alpha^s \geq 0$
β^d	Effect of π on demand	0.75	0.4	–	None
γ^d	Effect of r on demand	-1	0.4	–	$\gamma^d \leq 0$
ψ^y	Fed response to y	0.5	0.4	–	$\psi^y \geq 0$
ψ^π	Fed response to π	1.5	0.4	–	$\psi^\pi \geq 0$
	Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$				
ρ	Interest rate smoothing	0.5	0.2	–	$0 \leq \rho \leq 1$
	Priors for coefficients of C				
	Student t distribution with 3 degrees of freedom				
χ^m	Effect of z on u^m	0	0.2	–	None
	Priors for structural variances $\mathbf{D}\|\mathbf{A},\mathbf{C}$				
	Gamma distribution				
d_{ii}^{-1}	Reciprocal of variance	$1/\left(\mathbf{a}_i'\hat{\mathbf{S}}\mathbf{a}_i\right)$	$1/\left(\sqrt{2}\mathbf{a}_i'\hat{\mathbf{S}}\mathbf{a}_i\right)$	–	$d_{ii} > 0$
	Priors for lag coefficients $\mathbf{B}\|\mathbf{A},\mathbf{C},\mathbf{D}$				
	Normal distribution				
$\mathbf{b}_i$	Lagged coefficients of equation i	$\eta'\mathbf{a}_i$	$\sqrt{d_{ii}\mathbf{M}_i}$	–	None
In addition,					
b_{33}	Third element of monetary equation	ρ	$\sqrt{d_{33}/10}$	–	None

Table 2.8: **Prior distributions of the model with a relevant but invalid instrument** (M_2). This table shows the prior distribution of **A**, **D**, **B**, and **C** together with their hyper-parameters. For Student t distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

Parameter	**Meaning**	**Location**	**Scale**	**Skew**	**Sign restriction**
		Priors for contemporaneous coefficients of A			
		Student t distribution with 3 degrees of freedom			
α^s	Effect of π on supply	2	0.4	–	$\alpha^s \geq 0$
β^d	Effect of π on demand	0.75	0.4	–	None
γ^d	Effect of r on demand	-1	0.4	–	$\gamma^d \leq 0$
ψ^y	Fed response to y	0.5	0.4	–	$\psi^y \geq 0$
ψ^π	Fed response to π	1.5	0.4	–	$\psi^\pi \geq 0$
		Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$			
ρ	Interest rate smoothing	0.5	0.2	–	$0 \leq \rho \leq 1$
		Priors for coefficients of C			
		Student t distribution with 3 degrees of freedom			
χ^s	Effect of z on u^s	0	0.2	–	None
χ^d	Effect of z on u^d	0	0.2	–	None
χ^m	Effect of z on u^m	0	0.2	–	None
		Priors for structural variances $\mathbf{D}\vert\mathbf{A},\mathbf{C}$			
		Gamma distribution			
d_{ii}^{-1}	Reciprocal of variance	$1/\left(\mathbf{a}_i'\hat{\mathbf{S}}\mathbf{a}_i)\right)$	$1/\left(\sqrt{2}\mathbf{a}_i'\hat{\mathbf{S}}\mathbf{a}_i\right)$	–	$d_{ii} > 0$
		Priors for lag coefficients $\mathbf{B}\vert\mathbf{A},\mathbf{C},\mathbf{D}$			
		Normal distribution			
$\mathbf{b}_i$	Lagged coefficients of equation i	$\eta'\mathbf{a}_i$	$\sqrt{d_{ii}\mathbf{M}_i}$	–	None
In addition,					
b_{33}	Third element of monetary equation	ρ	$\sqrt{d_{33}/10}$	–	None

Estimation results for the baseline model

Combining the prior and the likelihood, the baseline model M_0 is estimated by the modified Baumeister-Hamilton algorithm, which is summarized in the proposition 1 above.[12] Figure 2.2 displays the prior and posterior distributions of contemporaneous parameters of the matrix **A**. For each panel, the red curves represent the prior, and the blue histograms represent the posterior. The prior and posterior distributions of β^d and ψ^y are significantly different, while they are quite similar for the rest of the parameters, suggesting a lack of identification.

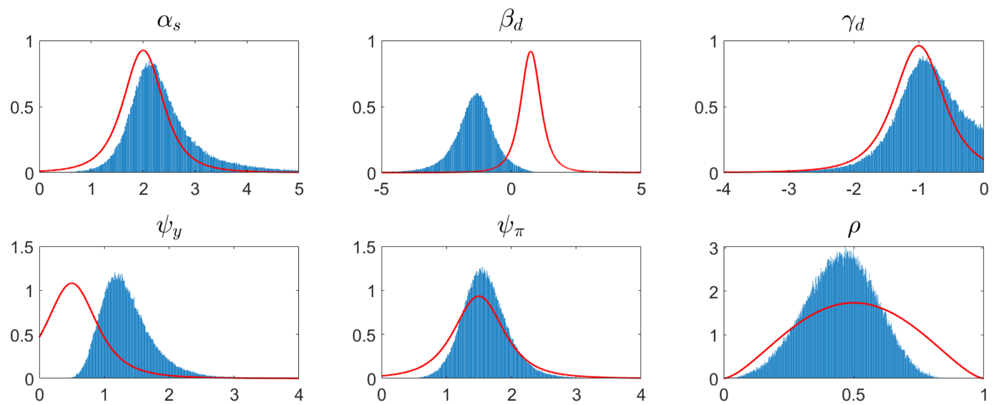

Figure 2.2: **Prior and posterior distributions of contemporaneous structural parameters of the baseline model (**M_0**)**. This figure shows the prior and posterior distributions of contemporaneous structural parameters (i.e. elements of **A** matrix) of the baseline model (M_0). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.

Thus, the data are informative about the effect of inflation on aggregate demand (β^d) and the effect of the output gap on the fed funds rates (ψ^y), but they are uninformative about other

[12]In all of the applications, I simulate 2,000,000 draws by the Metropolis-Hasting algorithm and discard the first 1,000,000 draws in order to ensure convergence of the Markov Chain.

structural parameters. In particular, the data suggest that higher inflation will reduce aggregate demand and the Fed has a stronger reaction to a positive output gap than we initially believe. The estimation results are mostly similar to those in Baumeister & Hamilton (2018), however, they found a smaller effect of inflation rate on aggregate demand and a smoother policy reaction function. The discrepancy can be explained by my different sample period that runs from 1954Q3 to 2008Q4 and includes the Volcker era of aggressive monetary policy and high inflation, while Baumeister & Hamilton (2018) restrict their sample to the Great Moderation period.

Bayesian investigation of instrument relevance

To formally investigate instrument relevance, I estimate model M_1 for each instrument and compute the Bayes factor with respect to M_0. Because the only difference between M_0 and M_1 is the use of the instrument in the latter model, the odds ratio is the indication for instrument relevance. Figures 2.3, 2.4, and 2.5 report the prior and posterior distributions of contemporaneous structural parameters together with the one that captures instrument relevance (χ^m) for the Romer-Romer, Sims-Zha, and Smets-Wouters instruments, respectively. In all those figures, the posterior distributions of the parameter χ^m are far away from zero and more concentrated than the prior, confirming that all three instruments are highly correlated with the monetary policy shock in the baseline model M_0.

Having a valid and relevant instrument also provides more identifying power to the models, and in this case, the instruments provide more information about the effect of the interest rate on aggregate demand (γ^d) and the inertia of the fed funds rate (ρ). In particular, the posterior distributions of γ^d are more concentrated near zero, suggesting a negative but small effect of the interest rate on aggregate demand. Also, the posterior distributions of ρ are more concentrated around higher values, implying that the Fed does have a desire to implement gradual change in its interest rate policy. This implication is different from that of the baseline model because most of the interest rate volatility in the Volcker era is now captured by the instruments.

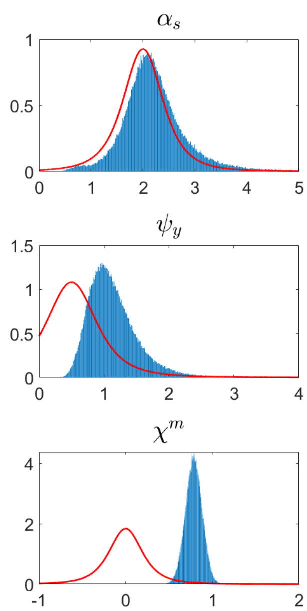

Figure 2.3: **Prior and posterior distributions of structural parameters of the alternative model (M_1) with the Romer-Romer instrument**. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $\mathbf{A}, \mathbf{C}$ matrices) of the alternative model where the instrument is valid (M_1). The instrument is the monetary policy shocks originally constructed by Romer & Romer (2004) and updated by Coibion, Gorodnichenko, Kueng & Silvia (2017a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.

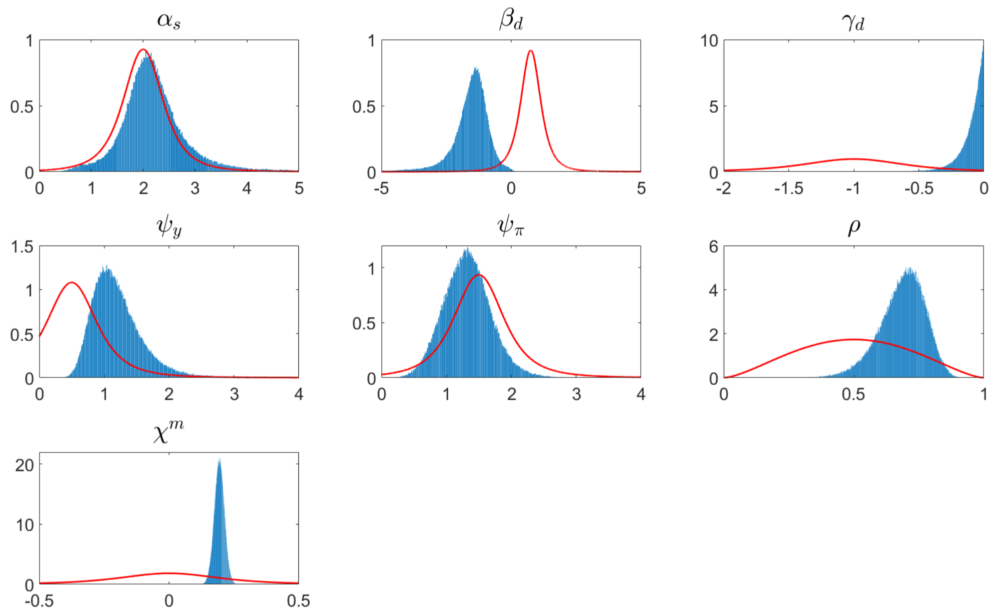

Figure 2.4: **Prior and posterior distributions of structural parameters of the alternative model (M_1) with the Sims-Zha instrument**. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $\mathbf{A}, \mathbf{C}$ matrices) of the alternative model where the instrument is not valid (M_1). The instruments are the monetary policy shocks estimated from Sims & Zha (2006)'s regime-switching SVAR and used in Stock & Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.

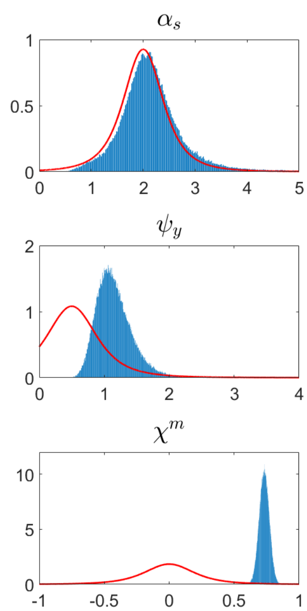

Figure 2.5: **Prior and posterior distributions of structural parameters of the alternative model (M_1) with the Smets-Wouters instrument**. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $\mathbf{A}, \mathbf{C}$ matrices) of the alternative model where the instrument is not valid (M_1). The instruments are interest rate shocks estimated from the medium-scale DSGE model of Smets & Wouters (2007) and used in Stock & Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.

Table 2.9 reports the marginal likelihood of the baseline model M_0, the alternative model M_1, and the test statistics. For each instrument, I calculate the marginal likelihood using four different methods. In the first two methods, I use Geweke (1999)'s proposal, and in the remaining two, I use Sims, Waggoner & Zha (2008)'s. Since both methods estimate the marginal likelihood from the truncated posterior distributions, I follow Herbst & Schorfheide (2015) and use two different levels of truncation, as represented by the value of the parameter τ. A value for τ of 0.5 means that we use 50% of the posterior draws and a value for τ of 0.9 means that we use 90% of the posterior draws. Table 2.9 shows that the estimates of the marginal likelihood and test statistics are robust to different methods.

Table 2.9: **Bayesian test for instrument relevance**. This table shows the marginal likelihood for the alternative and the baseline model, together with their difference. The baseline model, M_0, is the sign-restricted SVAR model estimated without using any external instrument, and the alternative model, M_1, is the sign-restricted SVAR model estimated with the instrument. Each instrument is entered one by one into the model. "Geweke" refers to the method proposed by Geweke (1999), and "SWZ" refers to the one proposed by Sims, Waggoner & Zha (2008). The parameter τ specifies the level of truncation of the posterior distribution. A value for τ of 0.5 means that 50% of the posterior draws are used, and a value for τ of 0.9 means that 90% of the posterior draws are used.

Test of instrument relevance-Bayesian approach			
	$\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T,M_0)$	$\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T,M_1)$	$2\times(\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T,M_1)-\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T,M_0))$
Romer-Romer instrument			
Geweke (tau=0.5)	-684	-662	44
Geweke (tau=0.9)	-684	-662	45
SWZ (tau=0.5)	-686	-664	45
SWZ (tau=0.9)	-685	-663	44
Sims-Zha instrument			
Geweke (tau=0.5)	-684	-650	68
Geweke (tau=0.9)	-684	-650	68
SWZ (tau=0.5)	-686	-653	67
SWZ (tau=0.9)	-685	-653	65
Smets-Wouters instrument			
Geweke (tau=0.5)	-684	-588	192
Geweke (tau=0.9)	-684	-588	191
SWZ (tau=0.5)	-686	-590	193
SWZ (tau=0.9)	-685	-590	191

How high does the Bayes factor have to be before we decide that model k is a better description of the data than model k'? Jeffreys (1967) emphasizes that the Bayes factor is a guide to decision making, thus its exact magnitude depends on the application. In some situations, decision makers might want to see a very large number of the Bayes factor before they reject the baseline model.[13] Table 2.10 reports a modified version of Jeffrey's guideline as presented by Kass & Raftery (1995). It describes different values of the test statistics, their corresponding Bayes factors, and the suggested interpretation. For ease of computation and comparison with the frequentist likelihood ratios, the statistics are calculated as two times the log transformation of the Bayes factor as shown in the first column of the table.

Table 2.10: **Jeffrey's criteria for model selection**, originally proposed by Jeffreys (1967). I use the modified version in Kass & Raftery (1995). The Bayes factor, $B_{k,k'}$, is the ratio of the marginal likelihood of the model M_k to the marginal likelihood of the model $M_{k'}$. When the prior belief is that the probabilities of the two models are equal, the Bayes factor is also their odds ratio.

Jeffrey's criteria		
2log($\mathbf{B}_{k,k'}$)	$\mathbf{B}_{k,k'}$	**Evidence against $\mathbf{M}_{k'}$**
0 to 2	1 to 3	Not worth more than a bare mentioning
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

According to Jeffreys' criteria from Table 2.10, there is very strong evidence that the instruments are relevant. For example, the test statistics for the Romer-Romer instrument are in the range of 44-45. It means that if a researcher's prior belief is that there is 50% chance the instrument is relevance, after seeing the data, she will revise her odd of the instrument being relevance to more than three billion to one. The test statistics for Sims-Zha's instrument and Smets-Wouters one have even higher values. These results are qualitatively similar to the

[13]"... (The Bayes factor) is not a physical magnitude. Its function is to grade the decisiveness of the evidence. It makes little difference to the null hypothesis whether the odds are 10 to 1 or 100 to 1 against it, and in practice no difference at all whether they are 10^4 or 10^{10} to 1 against it. In any case whatever alternative is most strongly supported will be set up as the hypothesis for use until further notice." *(Jeffreys, 1967, Appendix B)*

frequentist evidence in Table 2.3, namely all the instruments are strong but the Romer-Romer one is weaker than the others.

Bayesian investigation of instrument validity

To investigate instrument validity, I first estimate model M_2, which is the model where both relevance and validity of the instrument are in doubt. Figures 2.6, 2.7, and 2.8 report the prior and posterior distributions for the contemporaneous structural parameters and those associated with the Romer-Romer, Sims-Zha, and Smets-Wouters instrument, respectively. Posterior distributions for χ^m are similar to those of M_1, while those for χ^s and χ^d are different from their priors, suggesting that the data are also informative about those parameters. Specifically, posterior distributions of χ^s tend to concentrated around zero, and those of χ^d appear to have more mass in the positive support. Thus, the analysis suggests that the instruments are likely to be invalid. Consequently, they no longer provide as much identification power as before. Posterior distributions for ρ are still concentrated around the interval 0.4 to 0.8, however, these of γ^d are no longer skewed toward zero as much as those seen in the model M_1.

The frequentist tests of overidentifying restrictions, previously reported in Table 2.4 and Table 2.5, also suggest that at least one instrument is invalid, but they can't tell which instrument is valid and which one isn't. The Bayesian approach, on the other hand, allows us to make that determination. Table 2.11 reports the marginal likelihood for each model. The last column of Table 2.11 shows the test statistics for different instruments. Here, the tests do have the power to differentiate instruments based on their validity. For instance, the test statistics for the Romer-Romer instrument is between 7 and 9 across all four methods, suggesting that there is strong evidence that the Romer-Romer instrument is not valid. If a researcher has a prior belief that there is 50% chance the instrument is invalid, after seeing the data, she has to revise the odds of the instrument being invalid to about 50 to 1. Similarly, the test statistics for the Sims-Zha instrument is between 4 and 7, and thus there is positive to strong evidence that the instrument

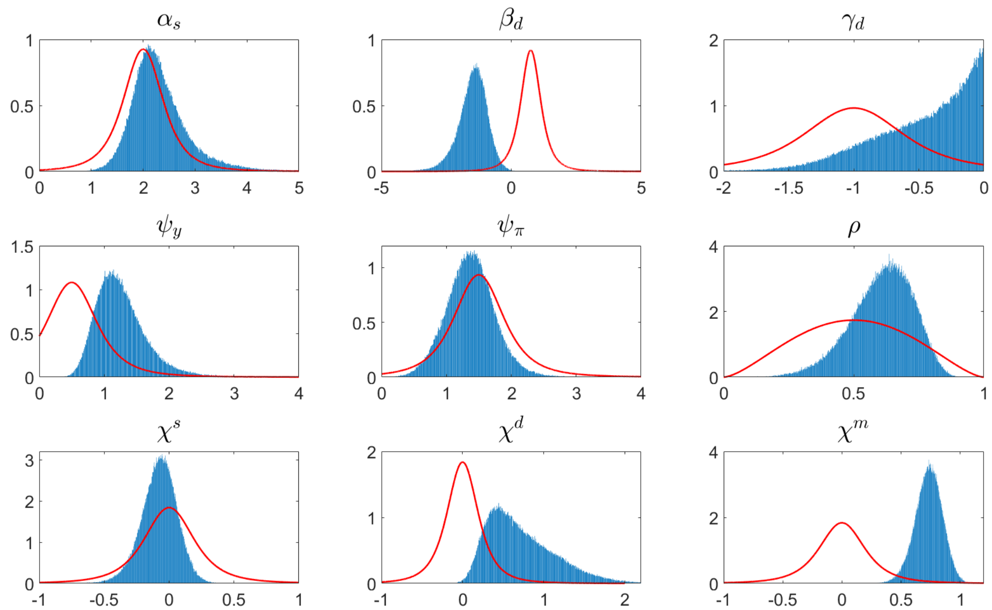

Figure 2.6: **Prior and posterior distributions of structural parameters of the alternative model (M_2) with the Romer-Romer instrument**. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $\mathbf{A}, \mathbf{C}$ matrices) of the alternative model where the instrument is not valid (M_2). The instruments are monetary policy shocks originally constructed by Romer & Romer (2004) and updated by Coibion, Gorodnichenko, Kueng & Silvia (2017a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.

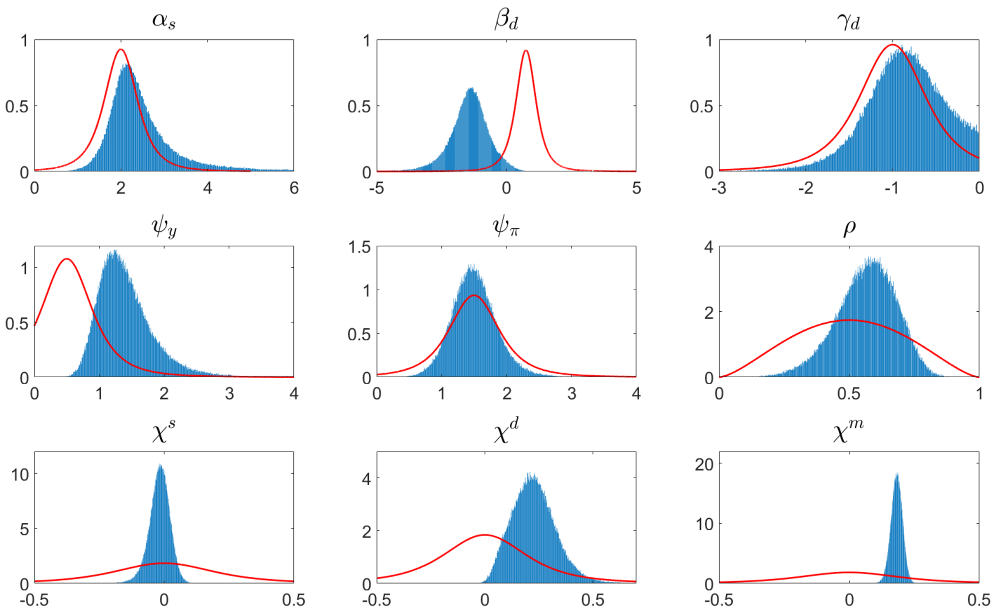

Figure 2.7: **Prior and posterior distributions of structural parameters of the alternative model (M_2) with the Sims-Zha instrument**. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $\mathbf{A}, \mathbf{C}$ matrices) of the alternative model where the instrument is not valid (M_2). The instruments are the monetary policy shocks estimated from Sims & Zha (2006)'s regime-switching SVAR and used in Stock & Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.

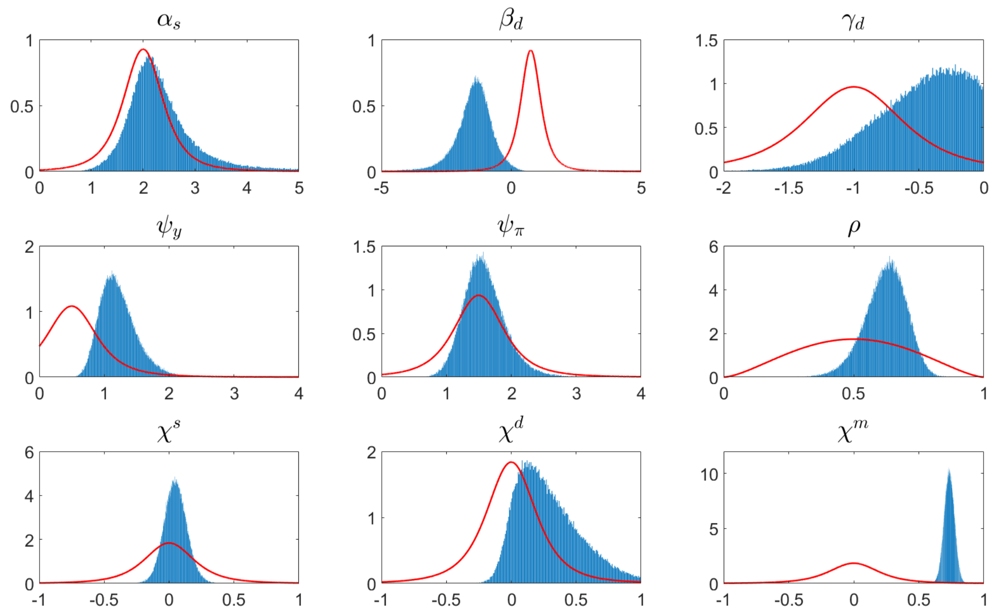

Figure 2.8: **Prior and posterior distributions of structural parameters of the alternative model (M_2) with the Smets-Wouters instrument**. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $\mathbf{A}, \mathbf{C}$ matrices) of the alternative model where the instrument is not valid (M_2). The instruments are interest rate shocks estimated from the medium-scale DSGE model of Smets & Wouters (2007) and used in Stock & Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.

is not valid. On the other hand, the test statistics for the Smets-Wouters instrument suggest that there is no significant evidence against the claim that the instrument is valid. Even if we use the most conservative number from table 2.11, the posterior odd against the claim of validity is only about 3 to 1. Therefore, the Bayesian approach suggests that the Romer-Romer and Sims-Zha instruments are invalid, whereas the Smets-Wouters instrument is valid.

Table 2.11: **Bayesian test for instrument validity**: This table shows the marginal likelihood for the alternative and the baseline model, together with their difference. The baseline model, M_1, is the sign-restricted SVAR model estimated with the external instrument entered the third equation, and the alternative model, M_2, is the sign-restricted SVAR model estimated with the instrument entered all three equations. The test is performed for one instrument at a time. ”Geweke” refers to the method proposed by Geweke (1999), and ”SWZ” refers to the one proposed by Sims, Waggoner & Zha (2008). The parameter τ specifies the level of truncation of the posterior distribution. A value for τ of 0.5 means that 50% of the posterior draws are used, and a value for τ of 0.9 means that 90% of the posterior draws are used.

Test of instrument validity-Bayesian approach			
	$\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T, M_1)$	$\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T, M_2)$	$2\times(\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T, M_2)-\mathbb{P}(\mathbf{Y}_T\|\mathbf{Z}_T, M_1))$
Romer-Romer instrument			
Geweke (tau=0.5)	-662	-659	7
Geweke (tau=0.9)	-662	-658	7
SWZ (tau=0.5)	-664	-660	9
SWZ (tau=0.9)	-663	-659	8
Sims-Zha instrument			
Geweke (tau=0.5)	-650	-649	4
Geweke (tau=0.9)	-650	-648	4
SWZ (tau=0.5)	-653	-650	6
SWZ (tau=0.9)	-653	-649	7
Smets-Wouters instrument			
Geweke (tau=0.5)	-588	-588	0
Geweke (tau=0.9)	-588	-588	1
SWZ (tau=0.5)	-590	-590	0
SWZ (tau=0.9)	-590	-589	1

2.3.4 Combining a relevant and valid instrument with sign restrictions

Replication of Romer and Romer (2004)'s specification

This section replicates the main regression in Romer & Romer (2004) to gain insight into the instruments' information content. One would expect that if the instruments have relevant information about the response of the output gap in a simple regression, it will be useful for identification in the sign-restricted SVAR. In their study, Romer & Romer (2004) use the following regression with monthly data

$$\Delta y_t = \alpha_0 + \sum_{k=1}^{11} \alpha_k D_k + \sum_{i=1}^{24} \beta_i \Delta y_{t-i} + \sum_{j=1}^{36} c_j S_{t-j} + e_t \tag{2.40}$$

where y is the log of industrial production, D is monthly dummies, and S is the Romer-Romer monetary policy shocks. In their specification, they assume that monetary policy instrument doesn't affect industrial production within a month. As pointed out in Hamilton (2017), equation (2.40) is equivalent to a two-variable VARX(36) with restrictions on the lag parameters. Because I have quarterly data, I will use a VARX(12) with restrictions on the lag parameters as follow

$$\begin{bmatrix} \Delta y_t \\ S_t \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix} + \sum_{k=1}^{8} \begin{bmatrix} \phi_{11}^{(k)} & \phi_{12}^{(k)} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{t-k} \\ S_{t-k} \end{bmatrix} + \sum_{k=9}^{12} \begin{bmatrix} 0 & \phi_{12}^{(k)} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{t-k} \\ S_{t-k} \end{bmatrix} + \sum_{k=1}^{3} \begin{bmatrix} 1 \\ 0 \end{bmatrix} D_k + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{2.41}$$

where y is the output gap, D is the quarterly dummies, and S is the monetary policy instrument, which is either that of Romer-Romer, Sims-Zha, or Smets-Wouters. For each instrument, I replace missing values with zero as done in Romer & Romer (2004) and estimate equation (2.41) with data from 1954Q3 to 2008Q4.

Figure 2.9 reports the SIRFs of output gap to one unit increase in the instruments together with their corresponding 95% confidence interval.[14] The patterns of response for all three

[14] Generally, one unit increase in the instrument doesn't translate to one percentage point increase in the fed funds

instruments are similar to that of Romer & Romer (2004), namely all SIRFs show a U-shape response of output gap to a contractionary monetary policy shock.[15] Thus, all three instruments deliver evidence that contractionary monetary policy shocks reduce the output gap.

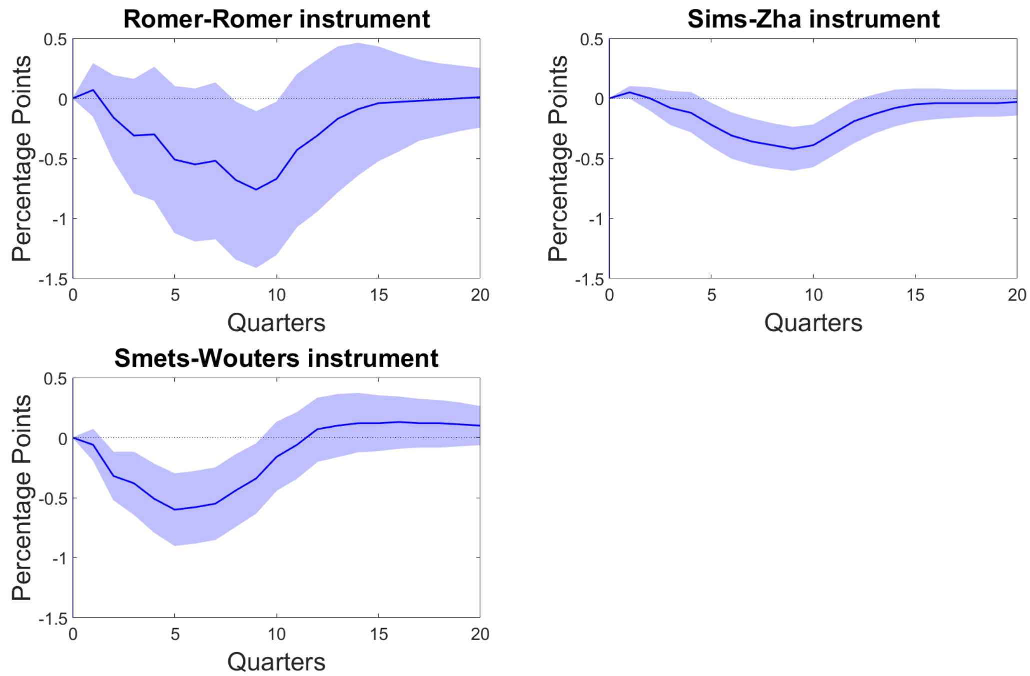

Figure 2.9: **Replication of Romer-Romer (2004)'s regression with data from 1954Q3 to 2008Q4**. This figure shows the impulse response function (IRF) of the output gap to one unit increase of the monetary policy instrument together with their 95 percent confidence interval. The three different monetary policy instruments are those of Romer & Romer (2004), Sims & Zha (2006), and Smets & Wouters (2007).

However, there are still some significant differences in the SIRFs in Figure 2.9. The SIRF estimated from the Smets-Wouters instrument suggests a faster response of output gap to interest rate shock than those estimated from the Romer-Romer and Sims-Zha instruments. In particular, the Smets-Wouters instrument implies a peak effect about one year after the shock, whereas those implied from Romer-Romer and Sims-Zha happen about two years after the shock. Moreover,

rate. Nakamura & Steinsson (2018) makes a similar observation.

[15]Instruments of monetary policy shocks and fed funds rates are positively correlated. Thus, an increase in the measures will be equivalent to a contractionary monetary policy shock.

although the magnitudes of those SIRFs are quite similar, their confidence intervals are not the same. Specifically, the SIRFs of Smets-Wouters and Sims-Zha instruments appear to be more precisely estimated than that of the Romer-Romer instrument.

In summary, although the empirical results from different instruments are qualitatively similar, there are still some significant differences regarding their practical implications. In light of previous evidence about instrument validity, I will only consider the Smets-Wouters instrument in the following section.[16]

Estimation results for the model with a relevant and valid instrument

In this section, I estimate model M_1 with the Smets-Wouters instrument from 1954Q3 to 2008Q4. The SIRFs are then compared with the baseline model M_0 where I don't incorporate the instrument. I report the median SIRFs together with their 95 percent credible set. Baumeister & Hamilton (2018) shows that this is an optimal inference if one assumes an absolute-value loss function regarding the SIRFs. Nonetheless, my credible set is not directly comparable to the above results from Romer-Romer's regression since their 95 percent confidence interval will tend to be larger than the 95 percent credible set in models that are only set-identified.

Since the main interest is in the effect of a monetary policy shock, I first focus on the results for the SIRFs of a contractionary monetary policy shock. Figure 2.5 shows the posteriors of the sign-restricted SVAR with the Smets-Wouters instrument. Besides χ^m, γ^d is also identified from the data since the posterior is highly concentrated around zero. To see the implication of this result, recall that the contemporaneous matrix is

[16] I also perform a similar analysis for the period of the Great Moderation. The U-shape pattern disappears. The evidence seems to depend mostly on the relationship between monetary policy and output before the Great Moderation period. The plots of the instruments suggest that most of their variations come from the Paul Volcker Era. These findings are similar to those of Coibion (2012), which found that the results of Romer & Romer (2004) are sensitive to the inclusion of the nonborrowed reserve targeting, 1979-1982. As Ramey (2016) points out, the consensus view is that monetary policy has been conducted more systematically in the recent period, thus pure monetary policy shocks are harder to identify and most of what we classify as "shocks" are just information effects from the Fed.

$$\mathbf{A}=\begin{bmatrix} 1 & -\alpha^S & 0 \\ 1 & -\beta^d & -\gamma^d \\ -(1-\rho)\psi^y & -(1-\rho)\psi^\pi & 1 \end{bmatrix}$$

which implies that the matrix for impact effect is

$$\mathbf{A^{-1}}=\frac{1}{|\mathbf{A}|}\begin{bmatrix} -\left[\beta^d+\gamma^d(1-\rho)\psi^\pi\right] & \alpha^s & \alpha^s\gamma^d \\ \gamma^d(1-\rho)\psi^y-1 & 1 & \gamma^d \\ -(1-\rho)\left(\psi^\pi+\beta^d\psi^y\right) & (1-\rho)(\psi^\pi+\alpha^s\psi^y) & \alpha^s-\beta^d \end{bmatrix}$$

where $|\mathbf{A}|=\alpha^s\left[1-\gamma^d(1-\rho)\psi^y\right]-\left[\beta^d+\gamma^d(1-\rho)\psi^\pi\right]$. When $\gamma^d=0$, this matrix becomes

$$\mathbf{A^{-1}}=\frac{1}{\alpha^s-\beta^d}\begin{bmatrix} -\beta^d & \alpha^s & 0 \\ -1 & 1 & 0 \\ -(1-\rho)\left(\psi^\pi+\beta^d\psi^y\right) & (1-\rho)(\psi^\pi+\alpha^s\psi^y) & \alpha^s-\beta^d \end{bmatrix}$$

In this case, a one unit increase of the monetary policy shock, u^m, increases the fed funds rate by one percentage point upon impact and have no contemporaneous effect on output and inflation. This implication is similar to the ”recursive assumptions” in the literature, which states that monetary policy shock doesn’t affect the output gap and inflation contemporaneously.

The above implication is verified from Figure 2.10, which shows the SIRFs of the output gap, inflation, and fed funds rate to a one unit increase of the monetary policy shock. This figure shows that the estimated SIRFs from the sign-restricted SVAR with external instrument tend to have a tight 95 percent credible sets, especially upon impact. Also, the impact effects on the output gap and inflation are closer to zero than the baseline model as suggested by the above derivation. The median SIRF shows that a contractionary monetary policy shock decreases the output gap for about two years after impact with the trough effect is around -0.3 percent. Then, the effect gets smaller and completely dies out about five years after. Thus, the sign-restricted SVAR with external instruments delivers a SIRF that has a similar shape to that of the Romer & Romer (2004)'s regression in Figure 2.9. This dynamic is also found in previous empirical literature but is different from that of Baumeister & Hamilton (2018)'s specification, which suggests a large, negative effect upon impact and dies out gradually over the next three years.[17]

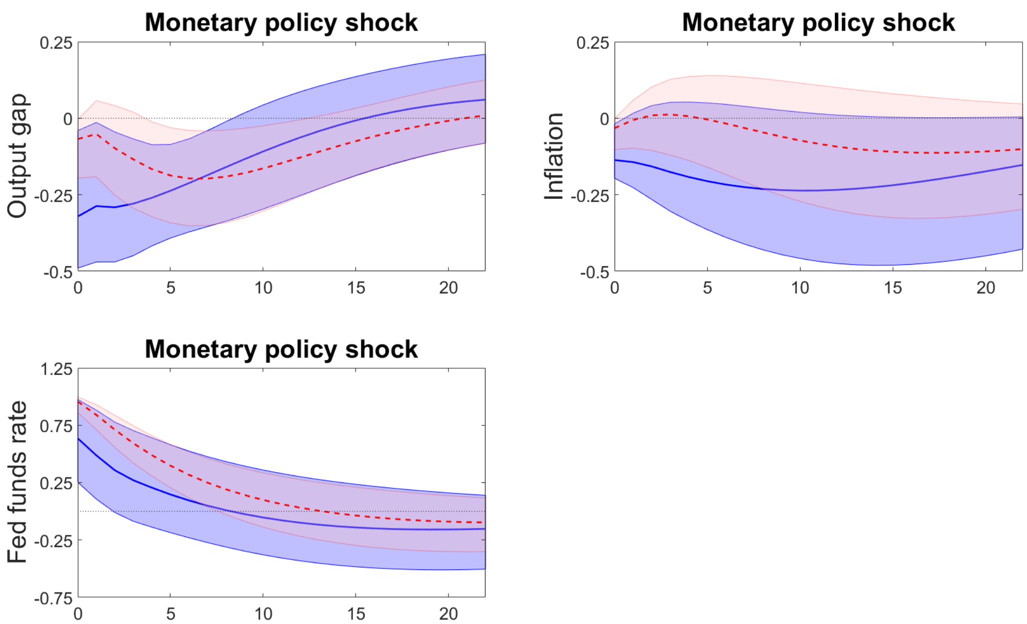

Figure 2.10: **SIRFs of monetary shocks estimated with the Smets-Wouters instrument**. This figure shows the estimated SIRFs of the sign-restricted SVAR augmented with the Smets-Wouters instrument together with that of the baseline model. The sample period is 1954Q3 to 2008Q4. The solid blue line is the median SIRF of the baseline model, and the dashed red line is the median SIRF for the model augmented with the instrument. The shaded blue region is the 95 percent credible set for the baseline model, and the shaded red region is the 95 percent credible set for the model augmented with the instrument.

Figure 2.11 shows the full result of all the SIRFs. Besides SIRFs for the monetary policy shock, I also show the SIRFs for the demand and supply shock. However, the results suggest that

[17]For example, Ramey (2016) studies the effect of monetary policy shock on industrial production using various specifications and found similar dynamics across all specification when the sample includes earlier observations in the 1970s. Her findings are shown in Figure 3.1, 3.2, and 3.3. She uses Christiano, Eichenbaum & Evans (1999)'s model, Coibion (2012)'s model, Jordà (2005)'s local projection method, proxy SVAR with Romer-Romer instrument, and proxy SVAR with high-frequency identification as in Gertler & Karadi (2015).

the instrument is not useful for shrinking down the identified set of these two shocks. The most significant difference is in the SIRF of inflation in response to the demand shock, in which the augmented model suggests a stronger response of inflation relative to that of the baseline model. This result is not surprising since the instrument is designed for monetary policy shock, and the previous identification analysis suggests that this instrument only provides information about the SIRFs of the monetary policy shock.

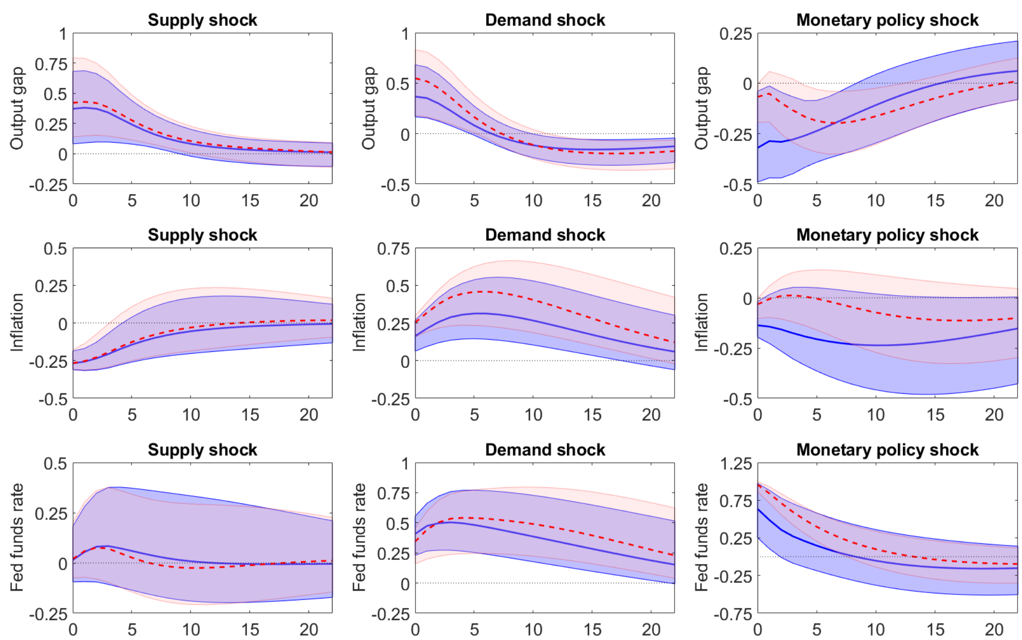

Figure 2.11: **SIRFs comparison between the sign-restricted SVAR augmented with Smets-Wouters instrument and the baseline model**. This figure shows the estimated SIRFs of the sign-restricted SVAR augmented with the Smets-Wouters instrument together with those of the baseline model. The sample period is 1954Q3 to 2008Q4. The solid blue line is the median SIRF of the baseline model, and the dashed red line is the median SIRF for the model augmented with the instrument. The shaded blue region is the 95 percent credible set for the baseline model, and the shaded red region is the 95 percent credible set for the model augmented with the instrument.

In summary, the analysis shows that my method can incorporate information from a relevant and valid instrument to improve inference of the dynamic causal effects. In this example, I use the instrument to identify the SIRFs of the monetary policy shocks, and I use the sign restrictions to identify the SIRFs of the supply and demand shocks. My results for the SIRFs of the monetary policy shock are qualitatively similar to those from the frequentist procedure in Romer & Romer (2004).

2.4 Conclusion

Identifying dynamic causal effects by external instruments is increasingly popular in empirical macroeconomics. However, its key identifying assumption, instrument validity, is either untestable in a just-identified model or fails to hold in some special cases when it is formally tested. Conventionally, researchers often defend that assumption by relying on *ad hoc* arguments to say that their instruments are ”plausibly exogenous”, but without a formal method to analyze instrument validity, they are simply advocating to replace incredible identification assumptions with untestable ones. The problem arises because existing econometric techniques cannot evaluate the validity assumption in a just-identified model nor can they say much about the validity of any particular instrument in an over-identified model. This paper aims to fill that gap.

I propose to use identifying assumptions from a sign-restricted SVAR as a benchmark to evaluate the underlying assumptions of the instrument. I first generalize the sign-restricted SVAR model developed by Baumeister & Hamilton (2015) to incorporate instrumental variables. Then, I show that my framework enables researchers to formally investigate the instrument relevance and validity assumptions by incorporating information from both theory and the empirical literature. In addition, my framework can combine information from relevant and valid instruments to improve inference in sign-restricted SVAR models. This latter contribution is of interest to practitioners who want to incorporate sign restrictions into a Proxy SVAR.

In the empirical application, I use the sign-restricted SVAR model of Baumeister & Hamilton (2018) to investigate the relevance and validity assumptions of three popular instruments for monetary shocks, proposed by Romer & Romer (2004), Sims & Zha (2006), and Smets &

Wouters (2007). My method not only replicates results from standard frequentist procedures, namely that the instruments are strong but all of them might not be valid, but it also differentiates valid from invalid instruments. In particular, I conclude that the Smets-Wouters instrument is valid, while the other two are not. Moreover, when the instrument is strong and valid, the framework delivers improved inferences for the SIRFs. I find that the SIRFs of the output gap follow a U-shape pattern in response to a contractionary monetary policy shock, and the effect reaches its maximum about one year after the shock then persists in the next four years. This finding is largely in line with the previous literature using external instruments, suggesting that my framework can incorporate information efficiently in practice.

Acknowledgement

Chapter 2, in full, is currently being prepared for submission for publication of the material. Lam Nguyen. The dissertation author was a primary author of this chapter.

Chapter 3

High-dimensional models and imperfect identifying information

Abstract

Current development in high-dimensional statistics fails to address the main interest of economists: causal inference with credible assumptions. I first review the literature on high-dimensional linear regression models and dynamic factor models. Then, I develop several new Bayesian numerical algorithms that combine the techniques in high-dimensional statistics with recent advances in dynamic causal inference. In particular, I discuss how to make causal statements from a high-dimensional structural model when researchers have doubts about identifying assumptions. Finally, I extend those algorithms to the case of Markov-switching models to accommodate nonlinearities in economic time series.

3.1 Introduction

Fueled by recent innovations in computing and data collection, empirical macroeconomists are increasingly interested in new methods that can work with large data set, also known as ”Big Data” (Hamilton (2006), Varian (2014), Ng (2017)). Examples of those data include disaggregated time series at the national level (McCracken & Ng (2016)), panel data across countries (Koop (2017)), and large data set from financial markets (Heaton, Polson & Witte (2016)). Economists are attracted to high-dimensional models with Big Data because they offer several advantages over traditional low-dimensional ones. First, they help reduce the omitted variable bias and model misspecifications in both forecasting and causal inference (Stock & Watson (2016)). Second, they validate economic theories that suggest most macroeconomic variables are driven by a few structural shocks (Sargent & Sims (1977)). Third, they improve the quality of real-time forecasts that aid the decision-making processes (Bańbura, Giannone, Modugno & Reichlin (2013)).

Despite the rise in popularity of Big Data, critics argue that their proponents have neglected the most important problem in economics: the identification problem. Economists always need identifying assumptions to convert observed correlations to causal relationship (Ramey (2016)). To that end, Big-Data practitioners often argue that they have solved that problem by using a large set of controlled variables, but critics, such as Charles Manski, are quick to point out that conditional exogeneity is hardly a convincing assumption (Tamer (2019)). Unfortunately, this serious matter receives little attention and most recent contributions in the macro time-series literature has instead focused on forecasting (Diebold, Ghysels, Mykland & Zhang (2019)).

The main contribution of this paper is to show how credible and robust causal inference is achieved by combining development in high-dimensional statistics with state-of-the-art identification strategies from economics. In particular, I develop several new Markov Chain Monte Carlo (MCMC) algorithms that could transparently incorporate both the identifying assumptions and researchers' doubt in a high-dimensional structural model, including both the structural vector autoregression (SVAR) and the structural dynamic factor model (SDFM). This identifying framework, developed by Baumeister & Hamilton (2015), has been shown to be a strict generalization of many conventional identification methods, and this paper is the first to implement it in the high-dimensional settings. Compared to the previous literature, such as Del Negro & Otrok (2007) and Amir-Ahmadi & Uhlig (2015), my procedure transparently incorporates priors and sign restrictions without using draws from the uniform Haar distribution to rotate the reduced-form covariance matrix, a practice known to impose implicit priors on the dynamic causal effects (Baumeister & Hamilton (2018), Baumeister & Hamilton (2019), and Watson (2019)).

The rest of this paper is organized as follows. Section 3.2 surveys the literature on high-dimensional linear regression models. Section 3.3 reviews the literature on dynamic factor models. Both sections focus on the Bayesian perspective. Then, section 3.4 presents several new MCMC algorithm to estimate high-dimensional structural models identified by sign restrictions, both linear and nonlinear ones. Section 3.5 concludes with directions for future research.

3.2 Review of high-dimensional linear regression models

One popular method to work with Big Data is to use a high-dimensional linear regression model. The literature is rising with many notable contributions, for instance De Mol, Giannone & Reichlin (2008), Bańbura, Giannone & Reichlin (2010), Giannone, Lenza & Primiceri (2015), Giannone, Lenza & Primiceri (2017), and Korobilis & Pettenuzzo (2019). The main idea is to develop priors that induce some forms of shrinkage on the coefficients with preferences given to those that are computationally convenient and satisfy certain optimal criteria.

Model description. To fix ideas, we will consider a typical high-dimensional linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{3.1}$$

$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2 \mathbf{I}\right) \tag{3.2}$$

where $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $(T \times 1)$ vectors of observed variables and innovations. $\mathbf{X}$ is an $(T \times k)$ vector of independent variables, $\boldsymbol{\theta}$ is an $(k \times 1)$ vector of coefficients, σ^2 is the variance. Since we are dealing with a high-dimensional linear regression model, we have $k >> T$.

A Bayesian's goal is to either analytically characterize or numerically simulate from the posterior distribution, which is proportional to the product of the prior and the likelihood by Bayes' rule. The choice of the Gaussian likelihood in (3.2) is generally uncontroversial and can be easily extended to the case of a fatter tail distribution such as the Student t distribution. Hence, most of the current research has focused on the choice of the priors for the high-dimensional vector of coefficients $\boldsymbol{\theta}$ in (3.1). I discuss next some of the popular choices for those priors.

Normal prior. One conventional prior for a linear regression model is the normal prior

$$\boldsymbol{\theta} \sim N\left(\bar{\theta}, \bar{\boldsymbol{\Sigma}}\right)$$

where $\bar{\theta}$ and $\bar{\boldsymbol{\Sigma}}$ are the prior means and covariance matrix. Researchers use those hyperparameters to incorporate additional information to improve the model's forecasting performance. For example, in a time-series setting, the so-called Minnesota prior of Litterman (1986) and Doan, Litterman & Sims (1984) shrinks the coefficients of the model toward random-walk behavior. Because this prior is a natural conjugate prior for the normal likelihood, the posterior distribution is analytically characterized and easy to simulate from. Bańbura, Giannone & Reichlin (2010) leverage this advantage and use a version of the Minnesota prior to estimate a large VAR with up to 131 variables. Nevertheless, in high-dimensional settings, researchers often don't have enough information about the relationship between various predictors to impose reasonable constraints and shrinkage. Hence, users of this prior often end up with models that have high posterior variances and deteriorated forecasting power. To address this problem, Giannone, Lenza & Primiceri (2015) suggest an empirical Bayes approach to estimate the shrinkage hyperparameters, but they have only tested their method with a medium-scale VAR with 22 variables. Thus, it is still unclear whether practitioners can successfully use this class of prior in high-dimensional settings.

Spike-and-slab prior. Mitchell & Beauchamp (1988) develop one of the earliest and most influential priors in the literature: the spike-and-slab prior. They suggest to use a mixture prior to represent the belief that the model parameter can either be zero (spike) or come from some proper distributions (slab). Assuming that the parameters are independent *a priori*, the prior can be written as

$$\boldsymbol{\theta} \sim \prod_{i=1}^{k} \left(p_i \times \delta_0 \left(\theta_i \right) + \left(1 - p_i \right) f \left(\theta_i \right) \right)$$

where p_i is the probability that the parameter θ_i is zero, $\delta_0(.)$ is the Dirac Delta function, and $f(.)$ can be an arbitrary probability distribution. Depending on the applications, researchers often vary on their choices of $f(.)$. For example, Mitchell & Beauchamp (1988) originally propose to use the uniform distribution, while Giannone, Lenza & Primiceri (2017) recently suggest to use the normal distribution. Despite the intuitive appeal of the prior, the Dirac Delta function makes it challenging to compute the posterior in high-dimensional models.

Stochastic search variable selection (SSVS) prior. George & McCulloch (1993) suggest the SSVS prior as a replacement for the computationally challenging spike-and-slab prior. Their idea is that a normal distribution centered at zero with a small variance is a good approximation for the Diract Delta function, and hence, they recommend to use a mixture of normal distributions: one with small and one with large variance. The prior can be written in the hierarchical form as

$$\boldsymbol{\theta} \sim \prod_{i=1}^{k} \left(\gamma_i \times N\left(0, \tau_i^2\right) + (1-\gamma_i)\, N\left(0, c_i^2 \tau_i^2\right)\right)$$

$$\gamma_i \sim \text{Bernoulli}\left(p_i\right)$$

This hierarchical representation of the prior facilitates the use of the Gibbs sampler, and thus gives it advantage over the spike-and-slab prior. Researchers mostly embrace the feature that variable selection is done automatically and they have extend the SSVS priors to a wide range of settings. For instance, Korobilis (2013) develops Gibbs samplers to estimate high-dimensional linear and nonlinear VAR with SSVS prior, and Jochmann, Koop & Strachan (2010) apply this prior to a high-dimensional VAR whose parameters subjected to unknown number of structural breaks. Nevertheless, the most serious setback of this approach is that there is no guarantee the sampler has visited every possible model specifications. Indeed, as the dimension (k) of the model increases, the number of possible models (i.e. 2^k) quickly exceeds the number of draws that can practically be done. This fact raises questions about attempt to use this prior to select a correct *structural VAR* without any additional economic information as in George, Sun & Ni (2008).

LASSO prior. Motivated by the popularity of the LASSO introduced by Tibshirani (1996), Park & Casella (2008) propose mixture distributions of normal and exponential priors. The resulting priors on the coefficients are Laplace distributions, which are equivalent to the LASSO's L_1 penalty on the likelihood function. The joint prior is written as

$$p\left(\boldsymbol{\theta}\right) = \prod_{i=1}^{k} p\left(\theta_i\right)$$

$$\theta_i | \tau_i^2, \sigma^2 \sim N\left(0, \tau_i^2 \sigma^2\right)$$

$$\tau_i^2 | \lambda \sim \exp\left(\frac{\lambda^2}{2}\right)$$

Similarly to that of the SSVS prior, the posterior distribution of the LASSO prior can be easily simulated from a Gibbs sampler. In addition, the Bayesian approach automatically provides finite sample characterization of the parameters and other quantities of interest. Belmonte, Koop & Korobilis (2014) extend this prior to time-varying parameter models and find successes in forecasting EU Area's inflation. However, when the predictors are highly collinear, the LASSO tends to produce poor predictions, which are dominated by those of ridge regression even in the low-dimensional settings.

Elastic net prior. To deal with the shortcomings of the LASSO, Zou & Hastie (2005) introduce the elastic net method. The elastic net penalty is basically a combination of the LASSO L_1-penalty and the ridge-regression L_2-penalty. Although the elastic net is originally proposed as a frequentist method, its Bayesian counterpart and the corresponding Gibbs sampler have been developed by Li, Lin, *et al.* (2010). In particular, the elastic net prior is

$$\begin{gathered} p(\boldsymbol{\theta}) = \prod_{i=1}^{k} p(\theta_i) \\ \theta_i|\tau_i,\sigma^2 \sim N\left(0, \left(\frac{\lambda_2}{\sigma^2}\frac{\tau_i}{\tau_i - 1}\right)^{-1}\right) \\ \tau_i|\sigma^2 \sim TG\left(\frac{1}{2}, \frac{8\lambda_2\sigma^2}{\lambda_1^2}, (1,\infty)\right) \\ \sigma^2 \sim \frac{1}{\sigma^2} \end{gathered}$$

where λ_1 and λ_2 are the user-specified hyper-parameters, and $TG(.,.,(1,\infty))$ is the truncated Gamma distribution with support over $(1,\infty)$.

Horseshoe prior. Most shrinkage priors, such as the LASSO prior, aim to reduce the noise in the data by shrinking all coefficients to zero. However, by having only one hyperparameter that controls the amount of shrinkage, they often miss out on large signals. To strike a balance in this trade-off, Carvalho, Polson & Scott (2010) introduce a new prior that can efficiently separate the signal from the noise. This prior, also known as the horseshoe prior, is written as

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{k} p(\theta_i)$$
$$\theta_i | \lambda_i, \tau \sim N\left(0, \tau^2 \lambda_i^2\right)$$
$$\tau \sim C^+(0,1)$$
$$\lambda_i \sim C^+(0,1)$$

where $C^+(0,1)$ is the half-Cauchy prior, whose support is the positive real line, with location parameter 0 and scale parameter 1. The parameter τ controls the global shrinkage of all the coefficients toward zero, whereas λ_i allows any local signal from the i^{th} observation to be captured efficiently. A clear advantage of this prior is that there is no need for user-specified hyperparameters. Moreoever, as documented in Polson & Scott (2010), Polson & Sokolov (2019), Bhadra, Datta, Polson & Willard (2019), and Bhadra, Datta, Polson & Willard (2016), this state-of-the-art prior enjoys many optimal theoretical properties and outperforms the LASSO in many instances. Notwithstanding, its application in economics is still limited with mixed results. In terms of forecasting, Follett & Yu (2019) develops a Gibbs sampler to implement this prior in a high-dimensional VAR and find favorable results, while Cross, Hou & Poon (2020) argue that this prior is no better than the conventional Minnesota prior or SSVS prior.

3.2.1 Prior evaluation

With a plethora of possible prior choices, how could one choose between different priors? Prior choice is important in high-dimensional Bayesian statistics since it will have significant effects on the posterior. There are two concerns: first, the effect of priors might not be washed away even with infinite amount of data (Ghosal & Van der Vaart (2017)), and second, seemingly objective priors on the parameters might be highly informative on their nonlinear transformations which are the objects of interest (Efron (1973)). Current research has tried to come up with new priors that resolve those two problems.

To deal with the first concern, researchers have often evaluated priors based on frequentist criteria such as posterior consistency, the ability of the posterior to concentrate its mass in the true model asymptotically. For example, Sparks, Khare & Ghosh (2015) provide necessary and sufficient conditions for the posterior consistency under different g-priors, while Ghosh, Khare & Michailidis (2019) investigate regularity conditions for posterior consistency of a high-dimensional Bayesian VAR model. Polson & Scott (2010) review many popular global-local shrinkage priors, and they introduce a new asymptotic framework to facilitate their comparisons. Researchers have also used other criteria such as risk and rate of convergence (Castillo, Schmidt-Hieber & Van der Vaart (2015), van der Pas, Szabó & van der Vaart (2017)). Nevertheless, many Bayesians still find it unsatisfying to evaluate a Bayesian procedure with frequentist criteria (Consonni, Fouskakis, Liseo & Ntzoufras (2018)).

For economists who are interested in causal inference, the second concern is more serious because it implies that there might be implicit assumptions the users are not aware of. Ideally, one wants to rely on economic theory and empirical literature to guide one's prior, but it is hard to incorporate economic theory in high-dimensional settings. The issue of objective priors for low-dimensional models have been studied extensively and reviewed in many papers (Berger, Pericchi, Ghosh, Samanta, De Santis, Berger & Pericchi (2001)), however, there is still limited research for sparsity-induced priors in high dimensional models. Bernardo, Bayarri, Berger, Dawid, Heckerman, Smith & West (2003), Ghosh (2011), Consonni, Fouskakis, Liseo & Ntzoufras (2018), and Polson & Scott (2010) are some of the attempts to come up with different criteria for a "good" prior. Currently, the most promising approach seems to be the horseshoe prior of Carvalho, Chang, Lucas, Nevins, Wang & West (2008). The horseshoe prior not only does a good job at separating signal from the noise, it also has a special property known as "regular variation", which is preserved under nonlinear transformation (Bhadra, Datta, Polson & Willard (2016)).

Hence, the horseshoe priors can be said to be ”non-informative” in both the original parameters and their nonlinear transformations, which are the objects of interest.

3.2.2 Posterior computation

Given the choice of priors, how could one compute the posterior distributions? For most of the priors introduced previously, researchers can efficiently simulate from the posterior distributions with the help of the Gibbs sampler. However, computation is always a challenge in high-dimensional setting. In general, researchers need to use numerical methods such as MCMC, Sequential Monte Carlo (SMC), Approximate Bayesian Computation (ABC), variational inference, or some combinations of them. The disadvantage of these methods is that they are difficult to implement in large models, or in other words, they are not scalable.

To facilitate computation, current research tries to impose specific structure into either the models or the priors so that the computation can be simplified without adversely affecting the quality of the inference. The preference is either for priors that results in posteriors that can be characterized analytically or can be approximated by simpler distributions. For example, Bańbura, Giannone & Reichlin (2010) suggest to use the Minnesota priors because the posterior can be characterized analytically. Koop, Korobilis & Pettenuzzo (2019) survey many papers where researchers impose particular structure on the covariance matrices to simplify the computation. West (2013) advocates forgetting factors as a way to use natural-conjugate analysis in nonlinear models. The advantage of these proposals are that they are more computationally efficient, however, those methods can only accommodate specific kinds of priors. In addition, researchers should be aware that certain techniques might impose implicit causal orders among the variables. For example, the popular decouple/recouple technique of West (2020) in high-dimensional time-varying parameter models is basically equivalent to the ”recursive assumption” which is widely criticized in economics (Ramey (2016), Nakamura & Steinsson (2018), Bognanni (2018)).

3.2.3 Empirical findings

Empirically, Bayesian shrinkage methods are shown to perform just as well as their frequentist counterparts and competing methods. For example, Park & Casella (2008) and Li, Lin, *et al.* (2010) show that the Bayesian LASSO and the Bayesian Elastic Net have similar performance as their frequentist counterparts when they are applied to real data. Using economic data, Bańbura, Giannone & Reichlin (2010) shows that a large Bayesian VAR with the Minnesota prior outperforms the FAVAR in forcasting. De Mol, Giannone & Reichlin (2008) provide further empirical evidence to show that Bayesian shrinkage methods with ridge-regression and LASSO penalty perform similarly to the Principal Component analysis. Those authors argue that their results are due to the fact that most macroeconomic time series are collinear, and hence, a few time series are good enough to capture most variation in the data.

3.3 Review of dynamic factor models

One serious criticism against the use of sparse high-dimensional linear regression models in economics is that they lack economic foundation. Critics have pointed out that economic theories generally do not support the assumption of sparsity (Giannone, Lenza & Primiceri (2017), Stock & Watson (2018), Tamer (2019)). Instead, economic theories often imply that most economic variables are moved by a few common shocks (Sargent & Sims (1977)), and hence dynamic factor models (DFM) are usually preferred as a tool for Big Data analysis. Most important, DFMs can easily accommodate many identification strategies used in lower-dimensional VAR (Stock & Watson (2016)). Bai & Ng (2008), Stock & Watson (2011), and Stock & Watson (2016) already survey the literature from the frequentist perspectives comprehensively, thus I will focus on the Bayesian literature in this section.

3.3.1 Motivation for dynamic factor models

Dynamic factor models are introduced to the economic literature by Geweke (1977) and Sargent & Sims (1977). Its central idea is to use a factor structure to reduce the dimension of high-dimensional parameters, which could be either regression coefficients or elements of covariance matrices. Otrok & Whiteman (1998) provides the first Bayesian treatment in economics, and since then, economists have applied the technique to a wide variety of settings, including the study of the housing market (Del Negro & Otrok (2007)), coincident index for economic activity (Stock & Watson (1991)), international business cycle (Del Negro & Otrok (2008), Kose, Otrok & Whiteman (2003), Crucini, Kose & Otrok (2011)), and various financial applications (Aguilar & West (2000), Omori, Chib, Shephard & Nakajima (2007)). An important special case of DFM is the factor-augmented VAR (FAVAR) models developed by Bernanke, Boivin & Eliasz (2005). This class of model generates a large literature on its own with many interesting applications. For example, Amir-Ahmadi & Uhlig (2009) and Amir-Ahmadi & Uhlig (2015) use FAVAR models identified by sign restrictions to study the effect of monetary policy shocks.

Practitioners also prefer DFMs over other approaches because economic relations change over time and they can leverage the large, well-developed literature on nonlinear DFMs. For instance, Kim & Nelson (1999) develop algorithms to estimate Markov-switching state space model, while Aguilar & West (2000), Kastner, Frühwirth-Schnatter & Lopes (2014)) and Kastner, Frühwirth-Schnatter & Lopes (2017) examine models with stochastic volatility. Other innovations include dynamic graphical and matrix models (Carvalho & West (2007), Carvalho, West & Bernardo, 2007, Wang & West (2009)), dynamic matrix models for stochastic volatility (Fox & West (2011)), time-varying sparsity modelling (Nakajima & West (2012), Nakajima & West, 2013), and nonlinear dynamical system (Bonassi, You & West (2011)). West (2013) and Gamerman & Salazar (2013) provide a comprehensive review on the subject.

Why should economists pay attention to Bayesian analysis of dynamic factor models given the well-developed frequentist literature? The answer is that Bayesian methods (1) provide more accurate characterization of statistical uncertainty, (2) can be extended easily to incorporate nonlinearity, and (3) can incorporate many state-of-the-art identification strategy as shown in the next section. To elaborate on the first point, frequentist estimation of DFMs often follows a two-step process, and hence it is subject to the generated-regressor problem. On the other hand, Bayesian approaches simulate the joint distribution of both the factors and the parameters, thus providing better characterization of statistical uncertainty. Second, the two-step process cannot be extended to nonlinear models such as time-varying parameter models and Markov-switching models. In these settings, the factors estimated from the first step are no longer consistent, and it is unclear how to to extend the two-step process to accommodate for nonlinearity. It is not a problem for Bayesian analysis as can be seen in the large Bayesian literature in time-varying parameter models with stochastic volatility and Markov-switching state-space models.

3.3.2 Econometric issues of dynamic factor models

To fix idea, suppose the dynamic of the data can be summarized by the following state-space representation

$$\mathbf{Y}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\varepsilon}_t \tag{3.3}$$

$$\mathbf{F}_t = \mathbf{\Phi}\mathbf{F}_{t-1} + \mathbf{v}_t \tag{3.4}$$

$$\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{C}) \tag{3.5}$$

$$\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{\Omega}) \tag{3.6}$$

$$\mathbb{E}\left(\boldsymbol{\varepsilon}_t \boldsymbol{v}_s'\right) = \mathbf{0} \qquad \forall t, s \tag{3.7}$$

In the measurement equation (3.3), $\mathbf{Y}_t$ is a high-dimensional $(k \times 1)$ vector of observed macroeconomic time series. $\boldsymbol{\varepsilon}_t$ is a high-dimensional $(k \times 1)$ vector which represents either measurement errors or idiosyncratic shocks to each time series. $\mathbf{F}_t$ is a $(r \times 1)$ low-dimensional vector which might consists of either unobserved factors or some elements of the observed time series and unobserved factors. $\mathbf{\Lambda}$ is a $(k \times r)$ matrix of factor loading. And finally, $\mathbf{C}$ is a $(r \times r)$ covariance matrices. In the transition/state equation (3.4), $\mathbf{\Phi}$ is an $(r \times r)$ matrix that governs the dynamic of the factors, and $\mathbf{\Omega}$ is an $(r \times r)$ covariance matrix.

Given this model, two important questions in the literature are (1) how to choose the number of factor and how to estimate the model. The frequentist treatment of dynamic factor models and more generally state-space models are thoroughly reviewed in Hamilton (1994a), Hamilton (1994b), Durbin & Koopman (2012) and Stock & Watson (2016), thus I will focus on Bayesian methods.

First, from the Bayesian perspective, the most popular technique to determine number of factors would be to use Bayesian model comparisons between models with different factors. More advance techniques include the use of Reversible Jump MCMC (Green (1995)), bridge sampling (Meng & Wong (1996))), and variable selection priors (George & McCulloch (1993)). Applications of such techniques can be seen in Carvalho, Chang, Lucas, Nevins, Wang & West (2008) and Frühwirth-Schnatter & Lopes (2010).

Second, to estimate the model, most Bayesians use MCMC methods such as the Gibbs sampler. Gibbs sampling is convenient in this application because of two facts: (1) conditioning on the factors, both the measurement and transition equations are linear and can be dealt with easily using standard methods, and (2) conditioning on all model parameters, the distribution of the factors can be computed using the Kalman filter. Researchers often vary on how to do each step. For example, to sample the hidden factors, Carlin, Polson & Stoffer (1992) introduce single-move Gibbs sampling, while Carter & Kohn (1994) propose multi-move Gibbs sampling where they make use of the joint distribution of all the factors. Another problem with the Gibbs sampling in this kind of model is that the factors and the factor loadings are sometime highly correlated, which leads to an inefficient sampler. To deal with this problem, Simpson, Niemi & Roy (2017) use the insight from Yu & Meng (2011) and develop a technique to increase the efficiency of MCMC algorithm in factor models.

Finally, Bayesian methods can be extended easily to deal with nonlinearity. Although there are some sophisticated models where researchers have to use some computationally intensive methods such as SMC and ABC, MCMC is still sufficient for most models. For example, the Gibbs sampler for Markov-switching state space model of Kim, Shephard & Chib (1998) only has one additional step: first, we sample the discrete hidden states with the help of the Hamilton filter; then, conditioning on the hidden states, the rest of the unknown parameters and factors can be simulated as in the linear cases. MCMC methods for time-varying parameters models with stochastic volatility are discussed extensively in Chib, Nardari & Shephard (2002), Chib, Nardari & Shephard (2006), Frühwirth-Schnatter (2006), Shephard (1994).

3.4 Algorithms for sign-restricted models in a data-rich environment

Recent advances in high-dimensional statistics have focused on reduced-form models for the purpose of improving forecasting performance, but economists are mostly interested in causal inference. To go from a reduced-form models to a structural model, economists need identifying assumptions. Practitioners of Big Data's techniques often rely on the conditional exogeneity or recursive assumptions, but those are rarely convincing assumptions in economics.

In this section, I show how researchers can combine state-of-the-art identification strategies and recent advances in the high-dimensional statistic literature for the purpose of causal inference in a high-dimensional settings. The good news is that most of the identification strategy from economists can be extended to the case of high-dimensional settings with little modifications. The key insight is to bypass the reduced-form representation of the data and use Bayesian methods to work directly with the structural models as advocated by Baumeister & Hamilton (2015).

The most straightforward method to incorporate Big Data into a *structural model* is the two-step processes described by Stock & Watson (2016): first, we use Principal Component analysis to estimate some factors that summarize essential features of the data, then, we treat the estimated factors as observations and estimate the rest of the parameters. This method is justified because under certain regularity conditions, the estimated factor is shown to be a consistent estimates of the true factors. The advantage of this approach is that it can be implemented using existing algorithm, such as the Baumeister-Hamilton algorithm for sign-restricted SVAR models. The disadvantages are that we might mischaracterize the uncertainty because of our treatment of the factors as known, and it is unclear how to do it in the presence of structural instability.

To avoid those problems, I will describe several Bayesian algorithms that can jointly estimate both the factors and the parameters in the data-rich environment. These algorithms accurately represent the uncertainty in both the estimated factors and the parameters, and they can be extended easily to the cases of time-varying parameter models. First, I start with linear high-dimensional SVAR models, linear dynamic factor models, and then I proceed to their extensions to the case of Markov-switching models.

3.4.1 High-dimensional SVAR models with imperfect identifying information

Suppose that the dynamics of the data are summarized by a high-dimensional SVAR model

$$\mathbf{A}\mathbf{y}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{u}_t \tag{3.8}$$

$$\mathbf{u}_t \sim \text{i.i.d}\, N(\mathbf{0}, \mathbf{D}) \tag{3.9}$$

where $\mathbf{y}_t$ and $\mathbf{u}_t$ are $(n \times 1)$ vectors of observed variables of interest and structural shocks at time t, while $\mathbf{x}_{t-1}$ is a $(k \times 1)$ vector of independent variables. $\mathbf{A}$ is an $(n \times n)$ matrix that governs contemporaneous relationships between the variables of interest, $\mathbf{B}$ is an $(n \times k)$ matrix of coefficients, and $\mathbf{D}$ is an $(n \times n)$ diagonal covariance matrix of the structural shocks. Since we are dealing with a high-dimensional SVAR model, we have $k >> n$.

This structural model admits a reduced-form representation

$$\mathbf{y}_t = \mathbf{\Phi} \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{3.10}$$

$$\boldsymbol{\varepsilon}_t \sim \text{i.i.d}\, N\left(\mathbf{0}, \mathbf{\Omega}\right) \tag{3.11}$$

where $\mathbf{\Phi} = \mathbf{A}^{-1}\mathbf{B}$ is the matrix of reduced-form coefficients, $\mathbf{\Omega} = \mathbf{A}^{-1}\mathbf{D}\left(\mathbf{A}^{-1}\right)'$ is the reduced-form covariance matrix, and $\boldsymbol{\varepsilon}_t$ is a vector of reduced-form shock at time t.

MCMC algorithms for model estimation

The literature on high-dimensional statistics mostly develops methods to estimate the reduced-form representation as described in (3.10) and (3.11), whereas for the purpose of causal inference, economists are mostly interested in the structural model as described in (3.8) and (3.9). The key observation to estimate the structural form is that conditioning on $\mathbf{A}$, the rest of the model parameters can be simulated from methods already developed in the statistic literature. In particular, a general Gibbs sampling algorithm can proceed as follows.

MCMC Algorithm 1. We start the algorithm with random initial values $\left(\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \mathbf{D}^{(0)}\right)$. Suppose we are at the m^{th} iteration with parameters $\left(\mathbf{A}^{(m)}, \mathbf{B}^{(m)}, \mathbf{D}^{(m)}\right)$ and we want to draw parameters $\left(\mathbf{A}^{(m+1)}, \mathbf{B}^{(m+1)}, \mathbf{D}^{(m+1)}\right)$ in the $(m+1)^{th}$ step. We do so using the following steps

1. **Step 1**: Draw the structural parameters, $\mathbf{A}^{(m+1)}$, from the conditional posterior distribution $\mathbf{A}|\mathbf{B}, \mathbf{D}, \mathbf{Y}_T$.

2. **Step 2**: Draw the parameters of the covariance matrix, $\mathbf{D}^{(m+1)}$, from the conditional posterior distribution $\mathbf{D}|\mathbf{A}, \mathbf{B}, \mathbf{Y}_T$.

3. **Step 3**: Draw the parameters of the coefficient matrix, $\mathbf{B}^{(m+1)}$, from the conditional posterior distribution $\mathbf{B}|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T$.

In general, with an arbitrary prior for **A** and a Gaussian likelihood, step 1 can be implemented by any generic algorithm such as the Metropolis-Hasting algorithm and its variation. To facilitate computation in step 2, we could use a natural-conjugate prior, such as the inverse-Gamma distribution. Finally, after getting a draw for **A** and **D**, we could apply any technique developed in the previous literature for step 3.

With some special priors, the above algorithm can be simplified further to arrive at more efficient algorithm. The next section will describe two such priors and their corresponding algorithms, namely the Bayesian LASSO priors and the SSVS priors.

Bayesian Lasso

Priors. To regulate the parameter in the high-dimensional matrix **B**, we will introduce the vector of regulating parameters $\boldsymbol{\eta}$. In particular, the joint prior can be decomposed as

$$P(\mathbf{A},\mathbf{B},\mathbf{D},\boldsymbol{\eta}) = P(\boldsymbol{\eta})\,P(\mathbf{A}|\boldsymbol{\eta})\,P(\mathbf{D}|\mathbf{A},\boldsymbol{\eta})\,P(\mathbf{B}|\mathbf{A},\mathbf{D},\boldsymbol{\eta})$$

We will allow the prior for each element η_{ij} to be independent from each other and exponentially distributed: $P(\boldsymbol{\eta}) = \prod_{i=1}^{n}\prod_{j=1}^{k} p(\eta_{ij}) = \prod_{i=1}^{n}\prod_{j=1}^{k} \frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2}{2}\eta_{ij}^2\right)$. Prior for the contemporaneous parameters, $P(\mathbf{A}|\boldsymbol{\eta})$, can be arbitrary. Because the elements in the matrix **A** represents some economic quantities, they should not depend on $\boldsymbol{\eta}$. Thus, we could set $P(\mathbf{A}|\boldsymbol{\eta}) = P(\mathbf{A})$. There is also no reason why the prior for the variance matrix **D** has to depend on $\boldsymbol{\eta}$. Thus, we will let the parameters for the covariance matrix **D** to be independent from each other and use the standard Inverse-Gamma distribution. Hence, $P(\mathbf{D}|\mathbf{A},\boldsymbol{\eta}) = \prod_{i=1}^{n} p(d_{ii}|\mathbf{A})$ where $p(d_{ii}|\mathbf{A}) \sim$ Inverse-Gamma (κ_i,τ_i) for $i = 1,\ldots,n$. Finally, unlike Baumeister and Hamilton's papers, we will use independent priors for every elemennts of the matrix $\mathbf{B}$. In particular, we have

$$\begin{aligned} P(\mathbf{B}|\mathbf{A},\mathbf{D},\boldsymbol{\eta}) &= \prod_{i=1}^{n} p(\mathbf{b}_i|\mathbf{D},\mathbf{A},\boldsymbol{\eta}) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{k} p\left(b_{ij}|\mathbf{D},\mathbf{A},\boldsymbol{\eta}\right) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{k} N\left(m_{ij},d_{ii}\eta_{ij}^{2}\right) \end{aligned}$$

MCMC Algorithm 2. Given the prior and the likelihood, the posterior distribution can be simulated using the collapsed Gibbs sampler.We start the algorithm with random initial values $\left(\boldsymbol{\eta}^{(0)},\mathbf{A}^{(0)},\mathbf{B}^{(0)},\mathbf{D}^{(0)}\right)$. Suppose we are at the m^{th} iteration with parameters $\left(\boldsymbol{\eta}^{(m)},\mathbf{A}^{(m)},\mathbf{B}^{(m)},\mathbf{D}^{(m)}\right)$ and we want to draw parameters $\left(\boldsymbol{\eta}^{(m+1)},\mathbf{A}^{(m+1)},\mathbf{B}^{(m+1)},\mathbf{D}^{(m+1)}\right)$ in the $(m+1)^{th}$ step. We do so using the following steps

1. **Step 1:** Draw the regulating parameters, $\boldsymbol{\eta}$, from the conditional posterior distribution $\boldsymbol{\eta}|\mathbf{A},\mathbf{B},\mathbf{D},\mathbf{Y}_T$. In particular, we will draw each $\frac{1}{\eta_{ij}^{2}}$ from the Inverse-Gaussian $\left(\sqrt{\frac{\lambda^2 d_{ii}}{\left(b_{ij}-m_{ij}\right)^2}},\lambda^2\right)$ for $i=1,\ldots,n$ and $j=1,\ldots,k$.

2. **Step 2:** Draw the structural parameters, $\mathbf{A}$, from the conditional posterior distribution $\mathbf{A}|\boldsymbol{\eta},\mathbf{Y}_T$ by using the target distribution in the Baumeister-Hamilton's algorithm.

3. **Step 3:** Draw the parameters of the covariance matrix, $\mathbf{D}$, from $\mathbf{D}|\mathbf{A},\boldsymbol{\eta},\mathbf{Y}_T$ by using the posterior distributions in the Baumeister-Hamilton's algorithm.

4. **Step 4:** Draw the parameters of the coefficient matrix, $\mathbf{B}$, from the conditional posterior matrix $\mathbf{B}|\mathbf{A},\mathbf{D},\boldsymbol{\eta},\mathbf{Y}_T$. To simulate from the conditional posterior distribution of $\mathbf{B}$, we will

draw $\mathbf{b}_i$ from $N\left(\mathbf{m}_i^*, d_{ii}\mathbf{M}_i^*\right)$ where

$$\mathbf{m}_i^* = \left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i\right)$$

$$\mathbf{M}_i^* = \left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}$$

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{x}_{0i}' & \dots & \mathbf{x}_{T-1}' & \mathbf{P}_i' \end{bmatrix}$$

$$\tilde{\mathbf{Y}}_i = \begin{bmatrix} \mathbf{y}_1'\mathbf{a}_i & \dots & \mathbf{y}_T'\mathbf{x}_{T-1}' & \mathbf{P}_i'\mathbf{m}_i \end{bmatrix}$$

$$\mathbf{P}_i = \begin{bmatrix} \tau_{i1} & & & \\ & \tau_{i2} & & \\ & & \ddots & \\ & & & \tau_{ik} \end{bmatrix}$$

We will repeat step 1 to 4 for a large number of time (i.e. 2 millions), then we will discard the first half of the sample and use the second half for subsequent analysis.

Stochastic Search Variable Selection (SSVS)

Priors. To regulate the parameter in the high-dimensional matrix $\mathbf{B}$, we will introduce the vector of dummy variables $\boldsymbol{\gamma}$. In particular, the joint prior can be decomposed as

$$P(\mathbf{A},\mathbf{B},\mathbf{D},\boldsymbol{\gamma}) = P(\boldsymbol{\gamma})P(\mathbf{A}|\boldsymbol{\gamma})P(\mathbf{D}|\mathbf{A},\boldsymbol{\gamma})P(\mathbf{B}|\mathbf{A},\mathbf{D},\boldsymbol{\gamma})$$

We will allow the prior for each element γ_{ij} to be independent from each other and Bernoulli distributed: $P\left(\gamma_{ij}=1\right) = 1-P\left(\gamma_{ij}=0\right) = p_{ij}$. And the joint prior is $P(\boldsymbol{\gamma}) = \prod_{i=1}^n \prod_{j=1}^k p_{ij}^{\gamma_{ij}}\left(1-p_{ij}^{\gamma_{ij}}\right)$. The prior for the contemporaneous parameters, $P(\mathbf{A}|\boldsymbol{\gamma})$, can be arbitrary and we will set $P(\mathbf{A}|\boldsymbol{\gamma}) = P(\mathbf{A})$. The prior for the covariance matrix, $\mathbf{D}$, is $P(\mathbf{D}|\mathbf{A},\boldsymbol{\gamma}) = \prod_{i=1}^n p(d_{ii}|\mathbf{A})$ where $p(d_{ii}|\mathbf{A}) \sim$ Inverse-Gamma (κ_i,τ_i) for $i=1,\dots,n$. Finally, we will use independent priors for every elements of the matrix $\mathbf{B}$. In particular, we have

$$\begin{aligned} P(\mathbf{B}|\mathbf{A},\mathbf{D},\boldsymbol{\gamma}) &= \prod_{i=1}^{n} p(\mathbf{b}_i|\mathbf{D},\mathbf{A},\boldsymbol{\gamma}) \\ &= \prod_{i=1}^{n} p(\mathbf{b}_i|\mathbf{D},\mathbf{A},\boldsymbol{\gamma}) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{k} N\left(m_{ij}, d_{ii}a_{ij}^2\tau_{ij}^2\right) \end{aligned}$$

where $a_{ij} = 1$ if $\gamma_{ij} = 0$ and $a_{ij} = c_{ij}$ if $\gamma_{ij} = 1$. Basically, we have a mixture prior for b_{ij}, that is $b_{ij}|\gamma_{ij} \sim (1-\gamma_{ij}) N\left(m_{ij}, d_{ii}\tau_{ij}^2\right) + \gamma_{ij} N\left(m_{ij}, d_{ii}c_{ij}^2\tau_{ij}^2\right)$.

MCMC Algorithm 3. Given the prior and the Gaussian likelihood, the posterior distribution can be simulated using the collapsed Gibbs sampler. We start the algorithm with random initial values $\left(\boldsymbol{\gamma}^{(0)}, \mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \mathbf{D}^{(0)}\right)$. Suppose we are at the m^{th} iteration with parameters $\left(\boldsymbol{\gamma}^{(m)}, \mathbf{A}^{(m)}, \mathbf{B}^{(m)}, \mathbf{D}^{(m)}\right)$ and we want to draw parameters $\left(\boldsymbol{\gamma}^{(m+1)}, \mathbf{A}^{(m+1)}, \mathbf{B}^{(m+1)}, \mathbf{D}^{(m+1)}\right)$ in the $(m+1)^{th}$ step. We do so using the following steps

1. **Step 1:** Draw the dummy variables, $\boldsymbol{\gamma}$, from the conditional posterior distributions

   $\boldsymbol{\gamma}|\mathbf{A},\mathbf{B},\mathbf{D},\mathbf{Y}_T$. The key insight here is that the conditional posterior distributions of the regulating parameters do not depend of the data. In particular, we will draw each γ_{ij} from the Bernoulli $\left(\dfrac{a_{ij}}{a_{ij}+b_{ij}}\right)$ for $i = 1,\ldots,n$ and $j = 1,\ldots,k$ where

$$a_{ij} = p\left(b_{ij}|\mathbf{A},\mathbf{D},\gamma_{ij} = 1\right) p\left(\gamma_{ij} = 1\right)$$
$$b_{ij} = p\left(b_{ij}|\mathbf{A},\mathbf{D},\gamma_{ij} = 0\right) p\left(\gamma_{ij} = 0\right)$$

2. **Step 2, 3, 4:** Similar to Algorithm 2.

3.4.2 Dynamic factor models with imperfect identifying information

A DFM can be cast in a state-space form and estimated using the Gibbs sampler. This section provides a new algorithm to estimate a state-space model with sign restrictions. Suppose the dynamic of the data can be summarized by the following state-space model

$$\mathbf{Y}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\varepsilon}_t \tag{3.12}$$

$$\mathbf{A}\mathbf{F}_t = \mathbf{B}\mathbf{F}_{t-1} + \boldsymbol{u}_t \tag{3.13}$$

$$\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{C}) \tag{3.14}$$

$$\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{D}) \tag{3.15}$$

$$\mathbb{E}\left(\boldsymbol{\varepsilon}_t \boldsymbol{u}_s'\right) = \mathbf{0} \qquad \forall t, s \tag{3.16}$$

In the measurement equation (3.12), $\mathbf{Y}_t$ is a high-dimensional $(k \times 1)$ vector of observed macroeconomic time series. $\boldsymbol{\varepsilon}_t$ is a high-dimensional $(k \times 1)$ vector which represents either measurement errors or idiosyncratic shocks to each time series. $\mathbf{F}_t$ is a $(r \times 1)$ low-dimensional vector which might consists of either unobserved factors or some elements of the observed time series and unobserved factors. $\mathbf{\Lambda}$ is an $(k \times r)$ matrix of factor loading. And finally, $\mathbf{C}$ is an $(k \times k)$ covariance matrices.

In the transition/state equation (3.13), $\mathbf{A}$ is an $(r \times r)$ matrix that governs contemporaneous relationship between elements of $\mathbf{F}_t$, $\mathbf{B}$ are the $(r \times r)$ matrices of lag coefficients, and $\mathbf{D}$ is an $(r \times r)$ diagonal matrix that represents the variance of the structural shocks. Assumption (3.16) separates the measurement equation (3.12) and transition equation (3.13) into two different linear regression equations. Thus, if we know the factors, drawing the rest of the parameters will be easy.

The above model is a structural model because each equation and innovation in (3.13) has some economic interpretations. One special issue related to dynamic factor models is how to assign economic interpretation to the estimated factors. To that end, Stock & Watson (2018) recommends the name-factor normalization strategy. Basically, researchers will assign a variable or set of similar variables to be driven by a particular factor, and thus, allowing the factor to take on the economic meaning of those variables.

For computational purpose, we also work with the reduced-form equations of (3.13), which are achieved by multiplying $\mathbf{A}^{-1}$ to both sides of (3.13)

$$\mathbf{F}_t = \mathbf{\Phi}\mathbf{F}_{t-1} + \mathbf{v}_t \tag{3.17}$$

$$\mathbf{v}_t \sim N\left(\mathbf{0}, \mathbf{\Omega}\right) \tag{3.18}$$

where $\mathbf{\Phi} = \mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{\Omega} = \mathbf{A}^{-1}\mathbf{D}\left(\mathbf{A}^{-1}\right)'$

MCMC Algorithm 4. We start the algorithm with random initial values. Suppose we are at the m^{th} iteration with parameters $\left(\mathbf{F}^{(m)}, \mathbf{\Lambda}^{(m)}, \mathbf{C}^{(m)}, \mathbf{A}^{(m)}, \mathbf{B}^{(m)}, \mathbf{D}^{(m)}\right)$ and we want to draw $\left(\mathbf{F}^{(m+1)}, \mathbf{\Lambda}^{(m+1)}, \mathbf{C}^{(m+1)}, \mathbf{A}^{(m+1)}, \mathbf{B}^{(m+1)}, \mathbf{D}^{(m+1)}\right)$ in the $(m+1)^{th}$ step. We do so using the following steps

1. **Step 1**: Draw the structural parameters of the transition equation, $\mathbf{A}, \mathbf{B}, \mathbf{D}$, from the conditional posterior distribution, $\mathbf{A}, \mathbf{B}, \mathbf{D} | \mathbf{F}, \mathbf{\Lambda}, \mathbf{C}, \mathbf{Y}^T$, by the Baumeister and Hamilton's algorithm.

2. **Step 2**: Draw the parameters of the state equation $\mathbf{\Lambda}, \mathbf{C}$, from the conditional posterior distribution, $\mathbf{\Lambda}, \mathbf{C} | \mathbf{F}, \mathbf{A}, \mathbf{B}, \mathbf{D}, \mathbf{Y}^T$, by standard method because this is just a linear regression when the factors are known.

3. **Step 3**: Draw the factors $\mathbf{F}$ from the conditional posterior distribution, $\mathbf{F} | \mathbf{\Lambda}, \mathbf{C}, \mathbf{A}, \mathbf{B}, \mathbf{D}, \mathbf{Y}^T$, by using the Kalman filter. In particular, assuming that $\mathbf{A}$ is invertible. When we condition on all the model parameters, we can multiply $\mathbf{A}^{-1}$ to both sides of equation (3.13) to convert it to (3.17). And then, we can apply the Kalman filter to make one draw for $\mathbf{F}$.

The above cycle will be repeated a large number of times to ensure convergence, then half the simulated draws are discarded to remove the effect of the initial conditions. The remaining half will be thought of as draws from the true posterior distribution and used for subsequent analysis.

3.4.3 Extensions to time-varying parameters

Most macroeconomic time series exhibit recurring structural breaks. One common way to model those events is to allow the model parameters to follow a hidden Markov chain and switch between different regimes. It is straightforward to extend the above Gibbs sampling algorithms to this class of models. The key insight is that conditioning on all the parameters, the hidden states can be sampled efficiently by the multi-move Gibbs sampling and the Hamilton filter. In the next paragraphs, I will present the algorithm to estimate both the Markov-switching high-dimensional SVAR models and Markov-switching state-space models.

Markov-switching high-dimensional SVAR models

Suppose that the dynamics of the data are summarized by a N-state Markov-switching high-dimensional SVAR model

$$\mathbf{A}_{S_t}\mathbf{y}_t = \mathbf{B}_{S_t}\mathbf{x}_{t-1} + \mathbf{u}_t$$

$$\mathbf{u}_t|S_t \sim N\left(\mathbf{0}, \mathbf{D}_{S_t}\right)$$

$$P\left(S_1 = s\right) = \mu_s$$

$$P\left(S_t = s'|S_{t-1} = s\right) = p_{s',s}$$

where $\mathbf{y}_t$ and $\mathbf{u}_t$ are $(n \times 1)$ vector of observed variables and structural shocks at time t, while $\mathbf{x}_{t-1}$ is a $(k \times 1)$ vector of independent variables. $\mathbf{A}_{S_t}$ is an $(n \times n)$ matrix that governs contemporaneous relationship between observed variables in state S_t, $\mathbf{B}_{S_t}$ is an $(n \times k)$ matrix of coefficients in state S_t, and $\mathbf{D}_{S_t}$ is an $(n \times n)$ diagonal covariance matrix of the structural shocks in state S_t. Since we are dealing with a high-dimensional SVAR model, we have $k >> n$. Let $\boldsymbol{\mu}$ be the vector of (μ_s)

for $s=1,\ldots,N$, $\mathbf{P}$ be the matrix of $\left(p_{s',s}\right)$ for $s,s'=1,\ldots,N$, and $(\mathbf{S}_T)$ be the vector of hidden states. The key difference from the linear model is that the parameters of this model depend on a hidden state, $(\mathbf{S}_T)$, which follows a Markov chain with initial probability $\boldsymbol{\mu}$ and transition probability matrix $\mathbf{P}$.

Markov-switching state-space models

Suppose that the dynamics of the data are summarized by a N-state state-space model

$$
\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\Lambda}_{S_t}\mathbf{F}_t + \boldsymbol{\varepsilon}_t \\
\mathbf{A}_{S_t}\mathbf{F}_t &= \mathbf{B}_{S_t}\mathbf{F}_{t-1} + \boldsymbol{u}_t \\
\boldsymbol{\varepsilon}_t &\sim N\left(\mathbf{0}, \mathbf{C}_{S_t}\right) \\
\mathbf{u}_t &\sim N\left(\mathbf{0}, \mathbf{D}_{S_t}\right) \\
\mathbb{E}\left(\boldsymbol{\varepsilon}_t \boldsymbol{u}_s'\right) &= \mathbf{0} \qquad \forall t,s \\
P\left(S_1 = s\right) &= \mu_s \\
P\left(S_t = s' | S_{t-1} = s\right) &= p_{s',s}
\end{aligned}
$$

Similarly, the key difference between this model and its linear counterpart is that all the model parameters depend on a hidden Markov chain $(\mathbf{S}_T)$. The definitions and interpretations of these parameters remain the same.

MCMC algorithm to estimate Markov-switching models

The key technique to estimate those models is data augmentation: conditioning on the states, the model is linear and its parameter can be estimated by the algorithm for linear model; similarly, conditioning on the model parameters, the hidden states can be sampled by the multimove Gibbs sampler and the Hamilton filter. Let $\boldsymbol{\Delta}$ denote all the unknown parameters in all the states except for those parameters related to the Markov chain. The generic MCMC algorithm to estimate both models are described below.

MCMC Algorithm 5. We start the algorithm with random initial values, $\left(\boldsymbol{\mu}^{(0)}, \mathbf{P}^{(0)}, \boldsymbol{\Delta}^{(0)}, (\mathbf{S}_T)^{(0)}\right)$. Suppose we are at the m^{th} iteration with parameters $\left(\boldsymbol{\mu}^{(m)}, \mathbf{P}^{(m)}, \boldsymbol{\Delta}^{(m)}, (\mathbf{S}_T)^{(m)}\right)$ and we want to draw $\left(\boldsymbol{\mu}^{(m+1)}, \mathbf{P}^{(m+1)}, \boldsymbol{\Delta}^{(m+1)}, (\mathbf{S}_T)^{(m+1)}\right)$. We do so using the following steps

1. **Step 1**: Draw the structural parameters, $\boldsymbol{\Delta}^{(m+1)}$, from their conditional posterior distribution

   $\boldsymbol{\Delta}|\boldsymbol{\mu}^{(m+1)}, \mathbf{P}^{(m+1)}, (\mathbf{S}_T)^{(m)}, \mathbf{Y}_T$

2. **Step 2**: Draw the initial probabilities, $\boldsymbol{\mu}^{(m+1)}$, from their conditional posterior distribution

   $\boldsymbol{\mu}|\mathbf{G}^{(m+1)}, \mathbf{P}^{(m)}, \boldsymbol{\Delta}^{(m)}, (\mathbf{S}_T)^{(m)}, \mathbf{Y}_T$

   Draw the transition probabilities, $\mathbf{P}^{(m+1)}$, from their conditional posterior distribution

   $\mathbf{P}|\boldsymbol{\mu}^{(m+1)}, \boldsymbol{\Delta}^{(m)}, (\mathbf{S}_T)^{(m)}, \mathbf{Y}_T$

3. **Step 3**: Draw the hidden states,$(\mathbf{S}_T)^{(m+1)}$, from their conditional posterior distribution

   $\mathbf{S}_T|\boldsymbol{\mu}^{(m+1)}, \mathbf{P}^{(m+1)}, \boldsymbol{\Delta}^{(m+1)}, \mathbf{Y}_T$

4. **Step 4**: Randomly permute all parameters and hidden states using the permutation sampler.

The above cycle is repeated a large number of times to ensure convergence, then half the simulated draws are discarded to remove the effect of the initial conditions. The remaining half will be thought of as draws from the true posterior distribution and used for subsequent analysis.

Step 1 in the above algorithm can be implemented using any of the previous algorithm for linear models. Step 2, 3, and 4 are specifically related to the Markov-switching model. Generally, for step 2 and 3, we will use the Dirichlet distribution, which is the natural-conjugate distribution for the vector of probability, to facilitate the computation. Finally, step 4 is necessary to deal with the label-switching problem in this class of model and ensure convergence of the Gibbs sampler.

3.5 Conclusion

Big Data presents new exciting opportunities and challenges for economists, but recent innovations have mostly focused on forecasting and relied mainly on conditional exogeneity as a main assumption for causal inference. This paper contributes to the literature by developing several Gibbs sampling algorithms that can combine the latest advances in the statistic literature with state-of-the-art identification strategy from the economic literature. In particular, I show how researchers can estimate both a high-dimensional SVAR and a SDFM with imperfect identifying information, such as sign restrictions.

Going forward, there are many interesting venues for application. For example, economists are increasingly interested in the heterogeneous effects of macroeconomic shocks and inequality (Coibion, Gorodnichenko, Kueng & Silvia (2017a), Kaplan, Moll & Violante (2018)) or in the effect of shocks on some continuous distributions such as income distributions or yield curves (Chang, Chen & Schorfheide (2018), Kowal, Matteson & Ruppert (2017), Kowal, Matteson & Ruppert (2019), Kowal & Bourgeois (2020)). Researchers can use a DFM with sign restrictions as a robust and flexible framework to study those questions empirically.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Lam Nguyen. The dissertation author was a primary author of this chapter.

Appendix A

Chapter 1 Appendix

A.1 Hamilton filter

In this Appendix, I describe the forward-backward algorithm for regime-switching model from Hamilton (1989), also known as the Hamilton filter. A more detailed exposition can be found in Hamilton (1994b) and Hamilton (2016).

Let $\boldsymbol{\eta}_t$ be an $N \times 1$ vector whose s-th element is

$$\begin{aligned} f\left(\mathbf{w}_t | S_t = s, \boldsymbol{\Omega}_{t-1}; \theta_{\neq S_T}\right) = & \frac{1}{(2\pi)^{\frac{n}{2}}} \left|\mathbf{A}_s\right| \left|\mathbf{D}_s\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{A}_s \mathbf{y}_t - \mathbf{B}_s \mathbf{x}_{t-1}\right)' \mathbf{D}_s^{-1} \left(\mathbf{A}_s \mathbf{y}_t - \mathbf{B}_s \mathbf{x}_{t-1}\right)\right\} \\ & \times \Pi_{k=1}^{K} P\left(\mathbf{z}_t^{(k)} | \mathbf{G}^{(k)}, S_t = s\right) \end{aligned}$$

And define $\hat{\boldsymbol{\xi}}_{t|t}$ to be an $N \times 1$ vector whose s-th element is $P\left(S_t = s | \boldsymbol{\Omega}_{t-1}, \boldsymbol{\theta}_{\neq S_T}\right)$. The Hamilton filter is the following recursion

$$\begin{aligned} \hat{\boldsymbol{\xi}}_{t|t-1} &= \mathbf{P} \hat{\boldsymbol{\xi}}_{t-1|t-1} \\ \hat{\boldsymbol{\xi}}_{t|t} &= \frac{\left(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}{\mathbf{1}' \left(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t\right)} \end{aligned}$$

Each element of $\hat{\boldsymbol{\xi}}_{t|t}$ is equal to the normalized $g\left(S_t|\boldsymbol{\theta}_{\neq S_T},\boldsymbol{\Omega}_t\right)$, which is what we need for the multimove Gibbs sampler. Furthermore, the filter will also give us the value of the likelihood function conditioning on the parameters because

$$L(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f\left(\mathbf{w}_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta}\right) = \sum_{t=1}^{T} \mathbf{1}'\left(\hat{\boldsymbol{\xi}}_{t|t-1} \bigodot \boldsymbol{\eta}_t\right)$$

Finally, to start the Hamilton filter, we need to specify the initial value, $\hat{\boldsymbol{\xi}}_{1|0}$. It can be set to equal to the vector of unconditional probabilities or can be treated as a separate parameters that reflect prior belief or are estimated by MLE.

A.2 Bayes Factor and Marginal Likelihood Estimation

The Bayes factor is the ratio of two marginal likelihoods, which are high dimensional integrals. Let $\boldsymbol{\theta}$ denote all unknown parameters in our model, the marginal likelihood for model l is defined as

$$P\left(\mathbf{W}_T|M_l\right) = \int P\left(\mathbf{W}_T|\boldsymbol{\theta}_l, M_l\right) P\left(\boldsymbol{\theta}_l|M_l\right) \mathrm{d}\boldsymbol{\theta}_\mathrm{l} \tag{A.1}$$

where $P\left(\mathbf{W}_T|\boldsymbol{\theta}_l, M_l\right)$ is the likelihood of model M_k, and $P\left(\boldsymbol{\theta}_l|M_l\right)$ is our prior belief. Equation (A.1) makes clear that the marginal likelihood is a weighted-average of the likelihood function over the prior distribution, and hence it is sensitive to our prior choices.[1] Although prior sensitivity is a potential problem when the prior distribution is chosen for technical convenience, it should not be a problem when the prior is carefully constructed from both theory and empirical literature as done in this case .

[1]In contrast, the likelihood ratio is the ratio of two likelihood functions evaluated at their maximizers. Since our model is only set-identified, it is unclear what the asymptotic distribution of the likelihood ratio will turn out to be.

Computation of the marginal likelihood is a technically challenging problem due to the large dimension of $\boldsymbol{\theta}$. In this paper, I use the methods proposed by Geweke (1999) and Sims, Waggoner & Zha (2008). Because I calculate the marginal likelihood for each model separately, I will drop the notation M_l in our discussion without loss of generality. Let $f(\boldsymbol{\theta})$ be a known, multivariate density function of $\boldsymbol{\theta}$, a simple application of Bayes theorem reveals that:

$$\frac{1}{P(\mathbf{W}_T)} = \int \frac{f(\boldsymbol{\theta})}{P(\mathbf{W}_T|\boldsymbol{\theta})\,P(\boldsymbol{\theta})} P(\boldsymbol{\theta}|\mathbf{W}_T)\,\mathrm{d}\boldsymbol{\theta}$$

Therefore, let N_0 be the size of the posterior draws after discarding the burn-in sample, a natural estimator for the marginal likelihood from posterior draws would be

$$\hat{P}(\mathbf{W}_T) = \left[\frac{1}{N_0}\sum_{n_0=1}^{N_0} \frac{f(\boldsymbol{\theta}^{n_0})}{P(\mathbf{W}_T|\boldsymbol{\theta}^{n_0})\,P(\boldsymbol{\theta}^{n_0})}\right]^{-1} \tag{A.2}$$

The choice of $f(\boldsymbol{\theta})$ is important in the calculation of the marginal likelihood. If we choose $f(\boldsymbol{\theta})$ to be a prior distribution (i.e. $f(\boldsymbol{\theta}) = P(\boldsymbol{\theta})$), we will have a harmonic mean estimator

$$\hat{P}_{HMM}(\mathbf{W}_T) = \left[\frac{1}{N_0}\sum_{n_0=1}^{N_0} \frac{1}{P(\mathbf{W}_T|\boldsymbol{\theta}^{n_0})}\right]^{-1}$$

However, this estimator has infinite variance and numerically inefficient. Indeed, for the method to work well, $f(\boldsymbol{\theta})$ needs to be a good approximation of the posterior distributions and also has a thinner tail than the posterior kernel $P(\mathbf{W}_T|\boldsymbol{\theta})\,P(\boldsymbol{\theta})$ to ensure convergence of the Monte Carlo average. Geweke (1999) proposes to use truncated normal distributions, while Sims, Waggoner & Zha (2008) constructs a more sophisticated choice for $f(\boldsymbol{\theta})$. I describe those two choices below. My description closely follows that of Herbst & Schorfheide (2015).

A.2.1 Geweke (1999)

Geweke (1999) proposes to use a truncated normal distribution as $f(\boldsymbol{\theta})$. Denote $\bar{\boldsymbol{\theta}}$ and $\bar{V}_{\boldsymbol{\theta}}$ be the mean and covariance matrix of $\boldsymbol{\theta}$. Both of those quantities are numerically computed from the posterior distributions. Geweke's choice of $f(\boldsymbol{\theta})$ is

$$f_{Geweke}(\boldsymbol{\theta}) = \tau^{-1}(2\pi)^{-d/2}|\bar{V}_{\boldsymbol{\theta}}|^{-1/2}\exp\left[-0.5\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)'\bar{V}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)\right]$$
$$\times\mathbb{I}\left\{\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)'\bar{V}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)\leq F_{\chi_d^2}^{-1}(\tau)\right\}$$

which is simply a truncated normal. $F_{\chi_d^2}^{-1}(\tau)$ denotes the inverse of the chi-squared distribution with d degrees of freedom, and the degree of truncation is controlled by τ. The lower value of τ , the more outliers will be remove from the posterior draws. In my empirical application, I try two different values for τ: $\tau_1 = 0.5$ and $\tau_2 = 0.9$.

A.2.2 Sims, Waggoner, and Zha (2008)

One shortcoming of the Geweke's proposal is that the posterior distribution might be very different from the Gaussian distributions, which will lead to poor estimate of the marginal likelihood. Sims, Waggoner & Zha (2008) proposes a more sophisticated choice for $f(\boldsymbol{\theta})$. Denote the mode of the posterior distribution as $\hat{\boldsymbol{\theta}}$, we calculate an analog of the covariance matrix that is centered at the mode of the distribution

$$\hat{V}_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\theta}^i-\hat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}^i-\hat{\boldsymbol{\theta}}\right)'$$

Next, define

$$r(\boldsymbol{\theta}) = \sqrt{\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)'\hat{V}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)}$$

And let $r^i = r\left(\boldsymbol{\theta}^i\right)$, we construct $f(\boldsymbol{\theta})$ in four steps

Step 1: Construct a heavy-tailed univariate density $g(r)$

$$g(r) = \begin{cases} \dfrac{vr^{v-1}}{b^v - a^v} & \text{if r} \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

where
$v = \dfrac{\ln(0.1/0.9)}{\ln(c_{10}/c_{90})}$, $a = c_1$,and $b = \dfrac{c_{90}}{0.9^{1/v}}$
(c_1, c_{10}, c_{90} are the first, 10th, and 90th percentiles of the empirical distribution of $\left\{r^i\right\}_{i=1}^{N}$)

Step 2: Define a new density $\tilde{f}(r)$ as

$$\tilde{f}(r) = \frac{\Gamma(d/2)}{2\pi^{d/2}\left|V_{\hat{\boldsymbol{\theta}}}\right|^{1/2}} \frac{g(r)}{r^{d-1}}$$

where $\Gamma(.)$ is the Gamma function, d is the dimension of the parameter vector $\boldsymbol{\theta}$.

Step 3: Define a truncating function

$$\mathbb{I}(\boldsymbol{\theta}) = \mathbb{I}\left(\ln P(\mathbf{Y}_T|\boldsymbol{\theta}, \mathbf{Z}_T) P(\boldsymbol{\theta}) > L_{1-q}\right) \times \mathbb{I}(r(\boldsymbol{\theta}) \in [a,b])$$

Then, we approximate the probability that function equal to 1 by simulation as

$$\hat{\tau} = \hat{\mathbb{P}}(\mathbb{I}(\boldsymbol{\theta}) = 1) = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}\left(\boldsymbol{\theta}^j\right)$$

where $\boldsymbol{\theta}^j$ is i.i.d and $\boldsymbol{\theta}^j \sim \tilde{f}(\boldsymbol{\theta})$.

Step 4: Sims, Waggoner, and Zha's choice of $f(\boldsymbol{\theta})$ is

$$f_{SWZ}(\boldsymbol{\theta}) = \hat{\tau}^{-1}\tilde{f}\left(\sqrt{\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)' \hat{V}_{\boldsymbol{\theta}}^{-1} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)}\right) \mathbb{I}(\boldsymbol{\theta})$$

The shortcoming of this method is that it is computationally expensive due to the fact that we have to estimate $\hat{\tau}$ by simulation.

A.3 Data Description

The data in this study are publicly available and are downloaded from FRED. In particular, I use the following five quarterly time series

1. GDPPOT: Real Potential Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Not Seasonally Adjusted.

2. GDPC1: Real Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Seasonally Adjusted Annual Rate

3. DPCERD3Q086SBEA (DPCE, for short): Personal consumption expenditures (implicit price deflator), Index 2009=100, Quarterly, Seasonally Adjusted

4. FEDFUNDS: Effective Federal Funds Rate, Percent, Quarterly Average, Not Seasonally Adjusted

5. USREC: NBER based Recession Indicators for the United States from the Period following the Peak through the Trough, +1 or 0, Quarterly, Not Seasonally Adjusted

The sample period is 1954Q3 to 2007Q4 since data for Effective Federal Funds Rate starts later relative to the first three. Denote $y =$ output gap, $\pi =$ inflation, and $r=$ fed funds rate. The variables in the baseline VAR are calculated from those time series as follow :

1. $y_t = 100 \times [\ln(GDPC1_t) - \ln\ (GDPPOT_t)]$

2. $\pi_t = 100 \times [\ln(DPCE_t) - \ln(DPCE_{t-4})]$

3. $r_t = FEDFUNDS_t$

Basically, the output gap is the difference between real and potential output, inflation is Y/Y change in PCE deflator, and the fed funds rate is the raw data.

Appendix B

Chapter 2 Appendix

B.1 Motivation for using conditional likelihood

B.1.1 Bayesian estimation

This Appendix shows that we can use conditional likelihood to estimate the model as long as the prior belief about the distribution of the exogenous variables are independent from the belief about the structural parameters. Consider the system of two equations:

$$\mathbf{A}\mathbf{y}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{z}_t + \mathbf{w}_t$$

$$\mathbf{z}_t = \mathbf{v}_t$$

where $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{D})$ and $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V})$. $\mathbf{D}$ is a diagonal matrix and $\mathbf{V}$ is allowed to be non-diagonal. We can define $\mathbf{y}_t^* = (\mathbf{y}_t', \mathbf{z}_t')'$, $\mathbf{A}^* = \begin{bmatrix} \mathbf{A} & -\mathbf{C} \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix}$, $\mathbf{D}^* = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{bmatrix}$, $\mathbf{B}^* = \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix}$, and $\mathbf{x}_{t-1} = \left(\mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \ldots, \mathbf{y}_{t-m}', \mathbf{1}\right)'$. Then, the system can be rewritten as

$$\mathbf{A}^*\mathbf{y}_t = \mathbf{B}^*\mathbf{x}_{t-1} + \mathbf{u}_t^*$$

Because $\det|\mathbf{A}^*| = \det|\mathbf{A}|$, the full likelihood function will be

$$\begin{aligned}
P(\mathbf{Y}_T^*|\mathbf{A}^*,\mathbf{B}^*,\mathbf{D}^*) = & = (2\pi)^{-T(n+r)/2} \det|\mathbf{A}^*|^T \det|\mathbf{D}^*|^{-T/2} \\
& \times \exp\left[-(1/2)\sum_{t=1}^{T}(\mathbf{A}^*\mathbf{y}_t^* - \mathbf{B}^*\mathbf{x}_{t-1})'\mathbf{D}^{*-1}(\mathbf{A}^*\mathbf{y}_t^* - \mathbf{B}^*\mathbf{x}_{t-1})\right] \\
& = (2\pi)^{-Tn/2} \det|\mathbf{A}|^T \det|\mathbf{D}|^{-T/2} \\
& \times \exp\left[-(1/2)\sum_{t=1}^{T}(\mathbf{A}\mathbf{y}_t - \mathbf{C}\mathbf{z}_t - \mathbf{B}\mathbf{x}_{t-1})'\mathbf{D}^{-1}(\mathbf{A}\mathbf{y}_t - \mathbf{C}\mathbf{z}_t - \mathbf{B}\mathbf{x}_{t-1})\right] \\
& \times (2\pi)^{-Tr/2} \det|\mathbf{V}|^{-T/2} \exp\left[-(1/2)\sum_{t=1}^{T}\mathbf{z}_t'\mathbf{V}^{-1}\mathbf{z}_t\right] \\
& = f(\mathbf{Y}_T|\mathbf{A}^*,\mathbf{D}^*,\mathbf{B}^*,\mathbf{Z}_T)\, f(\mathbf{Z}_T|\mathbf{V})
\end{aligned}$$

If the priors on $(\mathbf{A}^*,\mathbf{B},\mathbf{D})$ are independent from those on $\mathbf{V}$, then full system Bayesian inference for the elements of $\mathbf{A}^*,\mathbf{B},\mathbf{D}$ will be numerically identical to that based on the conditional likelihood

$$\begin{aligned}
P(\mathbf{Y}_T|\mathbf{A}^*,\mathbf{D}^*,\mathbf{B}^*,\mathbf{Z}_T) = & \ (2\pi)^{-Tn/2} \det|\mathbf{A}|^T \det|\mathbf{D}|^{-T/2} \\
& \times \exp\left[-(1/2)\sum_{t=1}^{T}(\mathbf{A}\mathbf{y}_t - \mathbf{C}\mathbf{z}_t - \mathbf{B}\mathbf{x}_{t-1})'\mathbf{D}^{-1}(\mathbf{A}\mathbf{y}_t - \mathbf{C}\mathbf{z}_t - \mathbf{B}\mathbf{x}_{t-1})\right]
\end{aligned}$$

Thus, the expression separates into two independent problems, and the posterior for the elements of $\mathbf{A}^*,\mathbf{B},\mathbf{D}$ can be simulated by the Baumeister-Hamilton algorithm as described in the main text.

B.1.2 Bayesian model comparison

Following the above motivation, this section shows why we can safely ignore the exogenous variable, $\mathbf{z}$, in the calculation of the marginal likelihood. Consider two model k and k', the posterior odd ratio can be written as

$$\frac{P(M_k|\mathbf{Y}_T,\mathbf{Z}_T)}{P(M_{k'}|\mathbf{Y}_T,\mathbf{Z}_T)} = \frac{P(\mathbf{Y}_T|\mathbf{Z}_T,M_k)}{P(\mathbf{Y}_T|,\mathbf{Z}_T,M_{k'})}\frac{P(\mathbf{Z}_T|M_k)}{P(\mathbf{Z}_T|M_{k'})}\frac{P(M_k)}{P(M_{k'})}$$

where

$$P(\mathbf{Z}_T|M_k) = \int P(\mathbf{Z}_T|\mathbf{\Psi},M_k)\,P(\mathbf{\Psi}|M_k)\,\mathrm{d}\mathbf{\Psi}$$
$$P(\mathbf{Z}_T|M_{k'}) = \int P(\mathbf{Z}_T|\mathbf{\Psi},M_{k'})\,P(\mathbf{\Psi}|M_{k'})\,\mathrm{d}\mathbf{\Psi}$$

where $\mathbf{\Psi}$ governs the likelihood of $\mathbf{Z}_T$. Therefore, as long as $P(\mathbf{Z}_T|\mathbf{\Psi},M_k) = P(\mathbf{Z}_T|\mathbf{\Psi},M_{k'})$ and $P(\mathbf{\Psi}|M_k) = P(\mathbf{\Psi}|M_{k'})$ for every $\mathbf{\Psi}$ in the parameter space, the two marginal likelihood will be the same and the expression simplifies to

$$\frac{P(M_k|\mathbf{Y}_T,\mathbf{Z}_T)}{P(M_{k'}|\mathbf{Y}_T,\mathbf{Z}_T)} = \frac{P(\mathbf{Y}_T|\mathbf{Z}_T,M_k)}{P(\mathbf{Y}_T|,\mathbf{Z}_T,M_{k'})}\frac{P(M_k)}{P(M_{k'})}$$

B.2 Marginal Likelihood Estimation

This Appendix describes the Geweke (1999) and Sims & Zha (2006) strategy to estimate the marginal likelihood. As described in the main text, the estimator for the marginal likelihood for model M_k is

$$\hat{P}(\mathbf{Y}_T|\mathbf{Z}_T,M_k) = \left[\frac{1}{N_0}\sum_{n_0=1}^{N_0}\frac{f(\boldsymbol{\theta}^{n_0})}{P(\mathbf{Y}_T|\boldsymbol{\theta}^{n_0},\mathbf{Z}_T,M_k)\,P(\boldsymbol{\theta}^{n_0}|M_k)}\right]^{-1} \tag{B.1}$$

The choice of $f(\boldsymbol{\theta})$ is important in numerical calculation. Geweke (1999) and Sims & Zha

(2006) propose different choices for $f(\boldsymbol{\theta})$. The description below closely follows that of Herbst & Schorfheide (2015).

B.2.1 Geweke (1999)

Geweke (1999) proposes to use a truncated normal distribution as $f(\boldsymbol{\theta})$. Particularly, denote $\bar{\boldsymbol{\theta}}$ and $\bar{V}_{\boldsymbol{\theta}}$ be the mean and covariance matrix of $\boldsymbol{\theta}$. Both of those quantities are numerically computed from the posterior distributions. Then, Geweke's choice of $f(\boldsymbol{\theta})$ is

$$f_{Geweke}(\boldsymbol{\theta}) = \tau^{-1}(2\pi)^{-d/2}|\bar{V}_{\boldsymbol{\theta}}|^{-1/2}\exp\left[-0.5\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)'\bar{V}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)\right]$$
$$\times\mathbb{I}\left\{\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)'\bar{V}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)\leq F_{\chi_d^2}^{-1}(\tau)\right\}$$

which is simply a truncated normal. $F_{\chi_d^2}^{-1}(\tau)$ denotes the inverse of the chi-squared distribution with d degrees of freedom, and the degree of truncation is controlled by τ. The lower value of τ , the more outliers will be remove from the posterior draws.

B.2.2 Sims, Waggoner, and Zha (2008)

One shortcoming of the Geweke's proposal is that the posterior distribution might be very different from the Gaussian distributions, which will lead to poor estimate of the marginal likelihood. Sims, Waggoner & Zha (2008) proposes a more sophisticated choice for $f(\boldsymbol{\theta})$. Denote the mode of the posterior distribution as $\hat{\boldsymbol{\theta}}$, then we can calculate an analog of the covariance matrix that is centered at the mode of the distribution

$$\hat{V}_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\theta}^i-\hat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}^i-\hat{\boldsymbol{\theta}}\right)'$$

Next, define

$$r(\boldsymbol{\theta}) = \sqrt{\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)'\hat{V}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)}$$

And let $r^i = r\left(\boldsymbol{\theta}^i\right)$, then we can construct $f(\boldsymbol{\theta})$ in four steps

1. Construct a heavy-tailed univariate density $g(r)$

$$g(r) = \begin{cases} \dfrac{vr^{v-1}}{b^v - a^v} & \text{if r} \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

   where $v = \dfrac{\ln(0.1/0.9)}{\ln(c_{10}/c_{90})}$, $a = c_1$, and $b = \dfrac{c_{90}}{0.9^{1/v}}$.
   (c_1, c_{10}, c_{90} are the first, 10th, and 90th percentiles of the empirical distribution of $\left\{r^i\right\}_{i=1}^{N}$)

2. Define a new density $\tilde{f}(r)$ as

$$\tilde{f}(r) = \frac{\Gamma(d/2)}{2\pi^{d/2}\left|V_{\hat{\boldsymbol{\theta}}}\right|^{1/2}} \frac{g(r)}{r^{d-1}}$$

   where $\Gamma(.)$ is the Gamma function, d is the dimension of the parameter vector $\boldsymbol{\theta}$.

3. Define a truncating function

$$\mathbb{I}(\boldsymbol{\theta}) = \mathbb{I}\left(\ln P(\mathbf{Y}_T|\boldsymbol{\theta},\mathbf{Z}_T)P(\boldsymbol{\theta}) > L_{1-q}\right) \times \mathbb{I}(r(\boldsymbol{\theta}) \in [a,b])$$

   Then, we can approximate the probability that function equal to 1 by simulation as

$$\hat{\tau} = \hat{\mathbb{P}}(\mathbb{I}(\boldsymbol{\theta}) = 1) = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}\left(\boldsymbol{\theta}^j\right)$$

   where $\boldsymbol{\theta}^j$ is i.i.d and $\boldsymbol{\theta}^j \sim \tilde{f}(\boldsymbol{\theta})$.

4. Sims, Waggoner, and Zha's choice of $f(\boldsymbol{\theta})$ is

$$f_{SWZ}(\boldsymbol{\theta}) = \hat{\tau}^{-1}\tilde{f}\left(\sqrt{\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)'\hat{V}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)}\right)\mathbb{I}(\boldsymbol{\theta})$$

The shortcoming of this method is that it is quite computationally expensive because we have to estimate $\hat{\tau}$ by simulation.

B.3 Data Description

The data in this study are publicly available. Macroeconomic time series are downloaded from FRED, while monetary policy shocks are collected from Mark Watson and Yuriy Gorodnichenko's website . Particularly, I use four quarterly time series from FRED.

1. GDPPOT: Real Potential Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Not Seasonally Adjusted.

2. GDPC1: Real Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Seasonally Adjusted Annual Rate

3. DPCERD3Q086SBEA (DPCE, for short): Personal consumption expenditures (implicit price deflator), Index 2009=100, Quarterly, Seasonally Adjusted

4. FEDFUNDS: Effective Federal Funds Rate, Percent, Quarterly Average, Not Seasonally Adjusted

The sample period is from 1954Q3 to 2008Q4 since data for Effective Federal Funds Rate starts later relative to the first three. Denote y= output gap, π= inflation, and r= fed funds rate. The variables in the baseline VAR will be calculated as follows

1. $y_t = 100 \times [\ln(GDPC1_t) - \ln\ (GDPPOT_t)]$

2. $\pi_t = 100 \times [\ln(DPCE_t) - \ln(DPCE_{t-4})]$

3. $r_t = FEDFUNDS_t$

Basically, the output gap is the difference between real and potential output, inflation is Y/Y change in the PCE deflator, and the fed funds rate is just the raw data.

For the choice of monetary policy shocks, I start with the same set of instruments used in Stock & Watson (2012a). Those four shocks are

1. Romer-Romer shock: Described in Romer & Romer (2004) and downloaded from Yuri Gorodnichenko's website. They are residuals from the regression between shocks constructed by the narrative methods on the Fed's Greenbook forecasts of output and inflation. The orginal data from their paper are from 1969Q1 to 1996Q4. In this study, I use the updated version constructed by and used in Coibion, Gorodnichenko, Kueng & Silvia (2017b). I use the quarterly data from their spreadsheet which are available from 1969Q1 to 2008Q4.

2. Smets-Wouters' shocks: Described in Smets & Wouters (2007) and downloaded from Mark Watson's website. They are interest rate shocks as calculated from Smets and Wouters' DSGE model. The data from Stock & Watson (2012a) study are calculated by King & Watson (2012). The data are available from 1959Q1 to 2004Q4.

3. Sims-Zha's shocks: Described in Sims & Zha (2006) and downloaded from Mark Watson's website. The data are supplied by Tao Zha. They are shocks constructed from the VAR that includes Markov-Switching variances and no time-varying parameters. Following Stock & Watson (2012a), I convert them to quarterly data by taking the monthly average. The data are available from 1960Q1 to 2003Q1.

Bibliography

1. Aguilar, O. & West, M. Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics* **18,** 338–357 (2000).

2. Amir-Ahmadi, P. & Drautzburg, T. Identification through heterogeneity. *Working Paper* (2017).

3. Amir-Ahmadi, P. & Uhlig, H. Measuring the dynamic effects of monetary policy shocks: a Bayesian FAVAR approach with sign restrictions. *Working Paper* (2009).

4. Amir-Ahmadi, P. & Uhlig, H. Sign Restrictions in Bayesian FaVARs with an application to monetary policy shocks. *Working Paper* (2015).

5. Angrist, J. D., Jordà, Ò. & Kuersteiner, G. M. Semiparametric estimates of monetary policy effects: string theory revisited. *Journal of Business & Economic Statistics* **36,** 371–387 (2018).

6. Antolín-Díaz, J. & Rubio-Ramírez, J. F. Narrative Sign Restrictions for SVARs. *Working Paper* (2016).

7. Antolín-Díaz, J. & Rubio-Ramírez, J. F. Narrative sign restrictions for SVARs. *American Economic Review* **108,** 2802–29 (2018).

8. Arezki, R., Ramey, V. A. & Sheng, L. News shocks in open economies: Evidence from giant oil discoveries. *Quarterly Journal of Economics* **132,** 103–155 (2017).

9. Arias, J., Rubio-Ramírez, J. F. & Waggoner, D. F. Inference in Bayesian Proxy-SVARs. *FRB Atlanta Working Paper* (2018).

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

11. Arias, J. E., Rubio-Ramírez, J. F. & Waggoner, D. F. Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica* **86,** 685–720 (2018).

12. Auerbach, A. J. & Gorodnichenko, Y. in *Fiscal policy after the financial crisis* 63–98 (University of Chicago Press, 2012).

13. Auerbach, A. J. & Gorodnichenko, Y. Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy* **4,** 1–27 (2012).

14. Auerbach, A. J. & Gorodnichenko, Y. Output spillovers from fiscal policy. *American Economic Review* **103,** 141–46 (2013).

15. Bahaj, S. A. Systemic sovereign risk: macroeconomic implications in the euro area. *Centre For Macroeconomics Working Paper* (2014).

16. Bai, J. & Ng, S. Large dimensional factor analysis. *Foundations and Trends® in Econometrics* **3,** 89–163 (2008).

17. Ball, L. & Mankiw, N. G. Asymmetric price adjustment and economic fluctuations. *Economic Journal* **104,** 247–261 (1994).

18. Bańbura, M., Giannone, D., Modugno, M. & Reichlin, L. Now-casting and the real-time data flow. *Handbook of Economic Forecasting* **2,** 195–237 (2013).

19. Bańbura, M., Giannone, D. & Reichlin, L. Large Bayesian vector auto regressions. *Journal of Applied Econometrics* **25,** 71–92 (2010).

20. Barnichon, R., Matthes, C. & Ziegenbein, A. Theory Ahead of Measurement? Assessing the Nonlinear Effects of Financial Market Disruptions. *Working Paper* (2016).

21. Bartlett, M. S. Comment on "A Statistical Paradox" by DV Lindley. *Biometrika* **44,** 533–534 (1957).

22. Basmann, R. L. On finite sample distributions of generalized classical linear identifiability test statistics. *Journal of the American Statistical Association* **55,** 650–659 (1960).

23. Baumeister, C. & Hamilton, J. D. Inference in structural vector autoregressions when the identifying assumptions are not fully believed: Re-evaluating the role of monetary policy in economic fluctuations. *Journal of Monetary Economics* **100,** 48–65 (2018).

24. Baumeister, C. & Hamilton, J. D. Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information. *NBER Working Paper* (2014).

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

26. Baumeister, C. & Hamilton, J. D. Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review* **109,** 1873–1910 (2019).

27. Baumeister, C. & Peersman, G. The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. *Journal of Applied Econometrics* **28,** 1087–1109 (2013).

28. Belmonte, M. A., Koop, G. & Korobilis, D. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting* **33,** 80–94 (2014).

29. Belongia, M. T. & Ireland, P. N. A Classical View of the Business Cycle. *NBER Working Paper* (2019).

30. Benati, L. & Surico, P. VAR analysis and the Great Moderation. *American Economic Review* **99,** 1636–52 (2009).

31. Berger, J. O., Pericchi, L. R., Ghosh, J., Samanta, T., De Santis, F., Berger, J. & Pericchi, L. Objective Bayesian methods for model selection: introduction and comparison. *Lecture Notes-Monograph Series,* 135–207 (2001).

32. Bernanke, B. S., Boivin, J. & Eliasz, P. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* **120,** 387–422 (2005).

33. Bernanke, B. S., Gertler, M. & Gilchrist, S. The financial accelerator in a quantitative business cycle framework. *Handbook of Macroeconomics* **1,** 1341–1393 (1999).

34. Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D, Smith, A. & West, M. Bayesian factor regression models in the "large p, small n" paradigm. *Bayesian Statistics* **7,** 733–742 (2003).

35. Bhadra, A., Datta, J., Polson, N. G. & Willard, B. Default Bayesian analysis with global-local shrinkage priors. *Biometrika* **103,** 955–969 (2016).

36. Bhadra, A., Datta, J., Polson, N. G. & Willard, B. Lasso meets horseshoe: A survey. *Statistical Science* **34,** 405–427 (2019).

37. Bognanni, M. A Class of Time-Varying Parameter Structural VARs for Inference under Exact or Set Identification. *Working Paper* (2018).

38. Bonassi, F. V., You, L. & West, M. Bayesian learning from marginal data in bionetwork models. *Statistical applications in genetics and molecular biology* **10** (2011).

39. Boschen, J. F. & Mills, L. O. The relation between narrative and money market indicators of monetary policy. *Economic Inquiry* **33,** 24–44 (1995).

40. Braun, R. & Brüggemann, R. Identification of SVAR Models by Combining Sign Restrictions With External Instruments. *Working Paper* (2017).

41. Caldara, D. & Herbst, E. Monetary policy, real activity, and credit spreads: Evidence from bayesian proxy svars. *American Economic Journal: Macroeconomics* **11,** 157–92 (2019).

42. Canova, F. & De Nicolo, G. Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics* **49,** 1131–1159 (2002).

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

44. Carlin, B. P. & Louis, T. A. *Bayes and empirical Bayes methods for data analysis* (Chapman and Hall/CRC, 2010).

45. Carlin, B. P., Polson, N. G. & Stoffer, D. S. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association* **87,** 493–500 (1992).

46. Carter, C. K. & Kohn, R. On Gibbs sampling for state space models. *Biometrika* **81,** 541–553 (1994).

47. Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q. & West, M. High-dimensional sparse factor modeling: applications in gene expression genomics. *Journal of the American Statistical Association* **103,** 1438–1456 (2008).

48. Carvalho, C. M., Polson, N. G. & Scott, J. G. The horseshoe estimator for sparse signals. *Biometrika* **97,** 465–480 (2010).

49. Carvalho, C. M. & West, M. Dynamic matrix-variate graphical models. *Bayesian Analysis* **2,** 69–97 (2007).

50. Carvalho, C. M., West, M. & Bernardo, J. Dynamic matrix-variate graphical models-A synopsis. *Bayesian Statistics VIII,* 585–90 (2007).

51. Castillo, I., Schmidt-Hieber, J. & Van der Vaart, A. Bayesian linear regression with sparse priors. *Annals of Statistics* **43,** 1986–2018 (2015).

52. Celeux, G., Hurn, M. & Robert, C. P. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association* **95,** 957–970 (2000).

53. Chan, J. C. & Tobias, J. L. Priors and posterior computation in linear endogenous variable models with imperfect instruments. *Journal of Applied Econometrics* **30,** 650–674 (2015).

54. Chang, M., Chen, X. & Schorfheide, F. Heterogeneity and Aggregate Fluctuations. *Working Paper* (2018).

55. Chib, S. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* **75,** 79–97 (1996).

56. Chib, S., Nardari, F. & Shephard, N. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics* **134,** 341–371 (2006).

57. Chib, S., Nardari, F. & Shephard, N. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* **108,** 281–316 (2002).

58. Christiano, L. J., Eichenbaum, M. & Evans, C. L. Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics* **1,** 65–148 (1999).

59. Cogley, T. & Sargent, T. J. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics* **8,** 262–302 (2005).

60. Cogley, T. & Sargent, T. J. Evolving post-world war II US inflation dynamics. *NBER Macroeconomics Annual* **16,** 331–373 (2001).

61. Coibion, O. Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics* **4,** 1–32 (2012).

62. Coibion, O., Gorodnichenko, Y., Kueng, L. & Silvia, J. Innocent Bystanders? Monetary policy and inequality. *Journal of Monetary Economics* **88,** 70–89 (2017).

63. Coibion, O., Gorodnichenko, Y., Kueng, L. & Silvia, J. Innocent Bystanders? Monetary Policy and Inequality in the US. *Journal of Monetary Economics,* 70–88 (2017).

64. Conley, T. G., Hansen, C. B. & Rossi, P. E. Plausibly exogenous. *Review of Economics and Statistics* **94,** 260–272 (2012).

65. Consonni, G., Fouskakis, D., Liseo, B. & Ntzoufras, I. Prior distributions for objective Bayesian analysis. *Bayesian Analysis* **13,** 627–679 (2018).

66. Cover, J. P. Asymmetric effects of positive and negative money-supply shocks. *Quarterly Journal of Economics* **107,** 1261–1282 (1992).

67. Cross, J. L., Hou, C. & Poon, A. Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting* (2020).

68. Crucini, M. J., Kose, M. A. & Otrok, C. What are the driving forces of international business cycles? *Review of Economic Dynamics* **14,** 156–175 (2011).

69. De Mol, C., Giannone, D. & Reichlin, L. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics* **146,** 318–328 (2008).

70. Del Negro, M. & Otrok, C. 99 Luftballons: Monetary policy and the house price boom across US states. *Journal of Monetary Economics* **54,** 1962–1985 (2007).

71. Del Negro, M. & Otrok, C. Dynamic factor models with time-varying parameters: measuring changes in international business cycles. *FRB of New York Staff Report* (2008).

72. Diebold, F. X., Ghysels, E., Mykland, P. & Zhang, L. *Big data in dynamic predictive econometric modeling* 2019.

73. Doan, T., Litterman, R. & Sims, C. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* **3,** 1–100 (1984).

74. Drautzburg, T. A narrative approach to a fiscal DSGE model. *Working Paper* (2016).

75. Durbin, J. & Koopman, S. J. *Time series analysis by state space methods* (Oxford university press, 2012).

76. Efron, B. Discussion of "Marginalization Paradoxes in Bayesian and Structural Inference," by AP Dawid, M. Stone, and JV Zidek. *Journal of the Royal Statistical Society* (1973).

77. Efron, B., Gous, A., Kass, R., Datta, G. & Lahiri, P. Scales of evidence for model selection: Fisher versus Jeffreys. *Lecture Notes-Monograph Series,* 208–256 (2001).

78. Faust, J. *The robustness of identified VAR conclusions about money* in *Carnegie-Rochester Conference Series on Public Policy* **49** (1998), 207–244.

79. Fisher, J. D. & Peters, R. Using stock returns to identify government spending shocks. *Economic Journal* **120,** 414–436 (2010).

80. Follett, L. & Yu, C. Achieving parsimony in Bayesian vector autoregressions with the horseshoe prior. *Econometrics and Statistics* **11,** 130–144 (2019).

81. Fox, E. B. & West, M. Autoregressive models for variance matrices: Stationary inverse Wishart processes. *arXiv preprint arXiv:1107.5239* (2011).

82. Frühwirth-Schnatter, S. *Finite mixture and Markov switching models* (Springer Science & Business Media, 2006).

83. Frühwirth-Schnatter, S. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* **96,** 194–209 (2001).

84. Frühwirth-Schnatter, S. & Lopes, H. F. Parsimonious Bayesian factor analysis when the number of factors is unknown. *Unpublished Working Paper, Booth Business* (2010).

85. Gafarov, B. Identification in Dynamic Models Using Sign Restrictions. *Working Paper* (2014).

86. Gafarov, B., Meier, M. & Olea, J. M. Projection Inference for Set-Identified SVARs. *Working Paper* (2016).

87. Gafarov, B., Meier, M. & Olea, J. L. M. Delta-Method inference for a class of set-identified SVARs. *Journal of Econometrics* **203,** 316–327 (2018).

88. Gamerman, D. & Salazar, E. Hierarchical modelling in time series: the factor analytic approach. *Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian FM Smith,* 167–182 (2013).

89. Garcia, R. & Schaller, H. Are the effects of monetary policy asymmetric? *Economic Inquiry* **40,** 102–119 (2002).

90. Gelfand, A. E. & Smith, A. F. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85,** 398–409 (1990).

91. Geman, S. & Geman, D. in *Readings in Computer Vision* 564–584 (Elsevier, 1987).

92. George, E. I. & McCulloch, R. E. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88,** 881–889 (1993).

93. George, E. I., Sun, D. & Ni, S. Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* **142,** 553–580 (2008).

94. Gertler, M. & Karadi, P. Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics* **7,** 44–76 (2015).

95. Geweke, J. *Contemporary Bayesian econometrics and statistics* (John Wiley & Sons, 2005).

96. Geweke, J. Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis* **51,** 3529–3550 (2007).

97. Geweke, J. The dynamic factor analysis of economic time series. *Latent variables in socio-economic models* (1977).

98. Geweke, J. Using simulation methods for Bayesian econometric models: inference, development, and communication. *Econometric Reviews* **18,** 1–73 (1999).

99. Ghosal, S. & Van der Vaart, A. *Fundamentals of nonparametric Bayesian inference* (Cambridge University Press, 2017).

100. Ghosh, M. Objective priors: An introduction for frequentists. *Statistical Science* **26,** 187–202 (2011).

101. Ghosh, S., Khare, K. & Michailidis, G. High-Dimensional Posterior Consistency in Bayesian Vector Autoregressive Models. *Journal of the American Statistical Association* **114,** 735–748 (2019).

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

103. Giacomini, R., Kitagawa, T. & Read, M. Robust Bayesian Inference in Proxy SVARs. *Working Paper* (2019).

104. Giacomini, R., Kitagawa, T. & Volpicella, A. Uncertain identification. *Working Paper* (2017).

105. Giannone, D., Lenza, M. & Primiceri, G. E. Economic predictions with big data: The illusion of sparsity. *FRB of New York Staff Report* (2017).

106. Giannone, D., Lenza, M. & Primiceri, G. E. Prior selection for vector autoregressions. *Review of Economics and Statistics* **97,** 436–451 (2015).

107. Granziera, E., Moon, H. R. & Schorfheide, F. Inference for VARs identified with sign restrictions. *Quantitative Economics* **9,** 1087–1121 (2018).

108. Green, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82,** 711–732 (1995).

109. Gürkaynak, R. S., Sack, B. & Swansonc, E. T. Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements. *International Journal of Central Banking* (2005).

110. Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society,* 357–384 (1989).

111. Hamilton, J. D. Computing power and the power of econometrics. *Medium Econometrische Toepassingen* **14,** 32–38 (2006).

112. Hamilton, J. D. *ECON 210D-Lecture 6: Event studies and high-frequency data* 2017.

113. Hamilton, J. D. in *Handbook of Macroeconomics* 163–201 (Elsevier, 2016).

114. Hamilton, J. D. State-space models. *Handbook of Econometrics* **4,** 3039–3080 (1994).

115. Hamilton, J. D. *Time series analysis* (Princeton New Jersey, 1994).

116. Hamilton, J. D. What is an oil shock? *Journal of Econometrics* **113,** 363–398 (2003).

117. Hamilton, J. D., Waggoner, D. F. & Zha, T. Normalization in econometrics. *Econometric Reviews* **26,** 221–252 (2007).

118. Heaton, J., Polson, N. G. & Witte, J. H. Deep learning in finance. *arXiv Preprint* (2016).

119. Herbst, E. P. & Schorfheide, F. *Bayesian estimation of DSGE models* (Princeton University Press, 2015).

120. Herwartz, H. & Lütkepohl, H. Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks. *Journal of Econometrics* **183,** 104–116 (2014).

121. Jasra, A., Holmes, C. C. & Stephens, D. A. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science,* 50–67 (2005).

122. Jefferson, P. N. Inference using qualitative and quantitative information with an application to monetary policy. *Economic Inquiry* **36,** 108–119 (1998).

123. Jeffreys, S. H. *Theory of Probability: 3d Ed* (Clarendon Press, 1967).

124. Jochmann, M., Koop, G. & Strachan, R. W. Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. *International Journal of Forecasting* **26,** 326–347 (2010).

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

126. Kaplan, G., Moll, B. & Violante, G. L. Monetary policy according to HANK. *American Economic Review* **108,** 697–743 (2018).

127. Karras, G. Why are the effects of money-supply shocks asymmetric? Convex aggregate supply or ”pushing on a string”? *Journal of Macroeconomics* **18,** 605–619 (1996).

128. Kass, R. E. & Raftery, A. E. Bayes factors. *Journal of the American Statistical Association* **90,** 773–795 (1995).

129. Kastner, G., Frühwirth-Schnatter, S. & Lopes, H. F. in *The Contribution of Young Researchers to Bayesian Statistics* 181–185 (Springer, 2014).

130. Kastner, G., Frühwirth-Schnatter, S. & Lopes, H. F. Efficient Bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics* **26,** 905–917 (2017).

131. Kaufmann, S. in *Advances in Markov-Switching Models* 137–157 (Springer, 2002).

132. Kaufmann, S. & Frühwirth-Schnatter, S. Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis* **23,** 425–458 (2002).

133. Kilian, L. Exogenous oil supply shocks: how big are they and how much do they matter for the US economy? *Review of Economics and Statistics* **90,** 216–240 (2008).

134. Kim, C.-J. & Nelson, C. R. State-space models with regime switching: classical and Gibbs-sampling approaches with applications. *MIT Press Books* **1** (1999).

135. Kim, S., Shephard, N. & Chib, S. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65,** 361–393 (1998).

136. King, R. G. & Watson, M. W. Inflation and Unit Labor Cost. *Journal of Money, Credit and Banking* **44,** 111–149 (2012).

137. Koop, G. Bayesian methods for empirical macroeconomics with big data. *Review of Economic Analysis* **9,** 33–56 (2017).

138. Koop, G., Korobilis, D. & Pettenuzzo, D. Bayesian compressed vector autoregressions. *Journal of Econometrics* **210,** 135–154 (2019).

139. Koop, G. & Potter, S. M. Bayes factors and nonlinearity: evidence from economic time series. *Journal of Econometrics* **88,** 251–281 (1999).

140. Korobilis, D. VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics* **28,** 204–230 (2013).

141. Korobilis, D. & Pettenuzzo, D. Adaptive hierarchical priors for high dimensional vector autoregressions. *Journal of Econometrics* (2019).

142. Kose, M. A., Otrok, C. & Whiteman, C. H. International business cycles: World, region, and country-specific factors. *American Economic Review* **93,** 1216–1239 (2003).

143. Kowal, D. R. & Bourgeois, D. C. Bayesian Function-on-Scalars Regression for High-Dimensional Data. *Journal of Computational and Graphical Statistics,* 1–26 (2020).

144. Kowal, D. R., Matteson, D. S. & Ruppert, D. A Bayesian multivariate functional dynamic linear model. *Journal of the American Statistical Association* **112,** 733–744 (2017).

145. Kowal, D. R., Matteson, D. S. & Ruppert, D. Functional autoregression for sparsely sampled data. *Journal of Business & Economic Statistics* **37,** 97–109 (2019).

146. Lanne, M. & Luoto, J. Data-driven inference on sign restrictions in bayesian structural vector autoregression. *CREATES Research Paper* **14** (2016).

147. Lanne, M., Lütkepohl, H. & Maciejowska, K. Structural vector autoregressions with Markov switching. *Journal of Economic Dynamics and Control* **34,** 121–131 (2010).

148. Li, Q., Lin, N., *et al.* The Bayesian elastic net. *Bayesian Analysis* **5,** 151–170 (2010).

149. Lindley, D. V. A statistical paradox. *Biometrika* **44,** 187–192 (1957).

150. Litterman, R. B. Forecasting with Bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics* **4,** 25–38 (1986).

151. Lo, M. C. & Piger, J. Is the response of output to monetary policy asymmetric? Evidence from a regime-switching coefficients model. *Journal of Money, Credit and Banking,* 865–886 (2005).

152. Ludvigson, S. C., Ma, S. & Ng, S. Shock Restricted Structural Vector-Autoregressions. *Working paper* (2017).

153. Lyu, Y. & Noh, E. Cyclical Variation in the Government Spending Multipliers: A Markov-switching SVAR Approach. *Working Paper* (2018).

154. McCracken, M. W. & Ng, S. FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* **34,** 574–589 (2016).

155. Meng, X.-L. & Wong, W. H. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica,* 831–860 (1996).

156. Mertens, K. & Ravn, M. O. The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review* **103,** 1212–47 (2013).

157. Mishkin, F. S. Is monetary policy effective during financial crises? *American Economic Review* **99,** 573–77 (2009).

158. Mitchell, T. J. & Beauchamp, J. J. Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **83,** 1023–1032 (1988).

159. Moon, H. R. & Schorfheide, F. Bayesian and frequentist inference in partially identified models. *Econometrica* **80,** 755–782 (2012).

160. Nakajima, J. & West, M. Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics* **31,** 151–164 (2013).

161. Nakajima, J. & West, M. Dynamic factor volatility modeling: A Bayesian latent threshold approach. *Journal of Financial Econometrics* **11,** 116–153 (2012).

162. Nakamura, E. & Steinsson, J. Identification in macroeconomics. *Journal of Economic Perspectives* **32,** 59–86 (2018).

163. Nevo, A. & Rosen, A. M. Identification with imperfect instruments. *Review of Economics and Statistics* **94,** 659–671 (2012).

164. Ng, S. Opportunities and Challenges: Lessons from Analyzing Terabytes of Scanner Data. *NBER Working Paper* (2017).

165. Noh, E. Uncovering an implicit restriction in Proxy-Structural VAR approach and its relaxation. *Working Paper* (2018).

166. Omori, Y., Chib, S., Shephard, N. & Nakajima, J. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* **140,** 425–449 (2007).

167. Otrok, C. & Whiteman, C. H. Bayesian leading indicators: measuring and predicting economic conditions in Iowa. *International Economic Review,* 997–1014 (1998).

168. Owyang, M. T. & Ramey, G. Regime switching and monetary policy measurement. *Journal of Monetary Economics* **51,** 1577–1597 (2004).

169. Owyang, M. T., Ramey, V. A. & Zubairy, S. Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data. *American Economic Review* **103,** 129–34 (2013).

170. Park, T. & Casella, G. The bayesian lasso. *Journal of the American Statistical Association* **103,** 681–686 (2008).

171. Paustian, M. Assessing sign restrictions. *The BE Journal of Macroeconomics* **7** (2007).

172. Peersman, G. & Smets, F. Are the effects of monetary policy in the euro area greater in recessions than in booms? *Monetary transmission in diverse economies,* 28–48 (2002).

173. Peersman, G. & Smets, F. The industry effects of monetary policy in the euro area. *Economic Journal* **115,** 319–342 (2005).

174. Piffer, M. Bayesian assessment of sign restrictions in VAR models. *Working Paper* (2017).

175. Plagborg-Møller, M. Bayesian inference on structural impulse response functions. *Quantitative Economics* **10,** 145–184 (2019).

176. Plagborg-Møller, M. & Wolf, C. K. Instrumental Variable Identification of Dynamic Variance Decompositions. *Working paper* (2017).

177. Poirier, D. J. Revising beliefs in nonidentified models. *Econometric Theory* **14,** 483–509 (1998).

178. Polson, N. G. & Scott, J. G. Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics* **9,** 501–538 (2010).

179. Polson, N. G. & Sokolov, V. Bayesian regularization: From Tikhonov to horseshoe. *Wiley Interdisciplinary Reviews: Computational Statistics* **11,** e1463 (2019).

180. Primiceri, G. E. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* **72,** 821–852 (2005).

181. Ramey, V. A. Identifying government spending shocks: it's all in the timing. *Quarterly Journal of Economics* **126,** 1–50 (2011).

182. Ramey, V. A. Macroeconomic shocks and their propagation. *Handbook of Macroeconomics* **2,** 71–162 (2016).

183. Ramey, V. A. & Zubairy, S. Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of Political Economy* **126,** 850–901 (2018).

184. Roberts, G. O. & Rosenthal, J. S. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* **16,** 351–367 (2001).

185. Romer, C. D. & Romer, D. H. A new measure of monetary shocks: Derivation and implications. *American Economic Review* **94,** 1055–1084 (2004).

186. Romer, C. D. & Romer, D. H. The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review* **100,** 763–801 (2010).

187. Rubio-Ramírez, J. F., Waggoner, D. F. & Zha, T. Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies* **77,** 665–696 (2010).

188. Rubio-Ramírez, J. F., Waggoner, D. F. & Zha, T. A. Markov-switching structural vector autoregressions: theory and application (2005).

189. Santoro, E., Petrella, I., Pfajfar, D. & Gaffeo, E. Loss aversion and the asymmetric transmission of monetary policy. *Journal of Monetary Economics* **68,** 19–36 (2014).

190. Sargan, J. D. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society,* 393–415 (1958).

191. Sargent, T. J. & Sims, C. A. Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research* **1,** 145–168 (1977).

192. Shephard, N. Partial non-Gaussian state space. *Biometrika* **81,** 115–131 (1994).

193. Simpson, M., Niemi, J. & Roy, V. Interweaving Markov chain Monte Carlo strategies for efficient estimation of dynamic linear models. *Journal of Computational and Graphical Statistics* **26,** 152–159 (2017).

194. Sims, C. A. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society,* 1–48 (1980).

195. Sims, C. A., Waggoner, D. F. & Zha, T. Methods for inference in large multiple-equation Markov-switching models. *Journal of Econometrics* **146,** 255–274 (2008).

196. Sims, C. A. & Zha, T. Were there regime switches in US monetary policy? *American Economic Review* **96,** 54–81 (2006).

197. Smets, F. & Wouters, R. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* **97,** 586–606 (2007).

198. Sparks, D. K., Khare, K. & Ghosh, M. Necessary and sufficient conditions for high-dimensional posterior consistency under *g*-priors. *Bayesian Analysis* **10,** 627–664 (2015).

199. Stephens, M. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 795–809 (2000).

200. Stock, J. H. Comment on Cogley and Sargent. *NBER Macroeconomics Annual* **2001,** 379–387 (2001).

201. Stock, J. H. & Watson, M. W. *A Probability Model of the Coincident Economic Indicators, in "Leading Economic Indicators: New Approaches and Forecasting Records", K. Lahiri and G. Moore, Eds* 1991.

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

203. Stock, J. H. & Watson, M. W. Dynamic factor models. *Oxford Handbooks Online* (2011).

204. Stock, J. H. & Watson, M. W. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics* **2,** 415–525 (2016).

205. Stock, J. H. & Watson, M. W. Has the business cycle changed and why? *NBER Macroeconomics Annual* **17,** 159–218 (2002).

206. Stock, J. H. & Watson, M. W. Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal* **128,** 917–948 (2018).

207. Stock, J. H. & Watson, M. W. *Introduction to econometrics: Global edition* (Pearson Education Boston, MA, 2012).

208. Tamer, E. Partial identification in econometrics. *Annual Review of Economics* **2,** 167–195 (2010).

209. Tamer, E. The ET Interview: Professor Charles Manski. *Econometric Theory* **35,** 233–294 (2019).

210. Tanner, M. A. & Wong, W. H. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82,** 528–540 (1987).

211. Tenreyro, S. & Thwaites, G. Pushing on a string: US monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics* **8,** 43–74 (2016).

212. Thoma, M. A. Subsample instability and asymmetries in money-income causality. *Journal of Econometrics* **64,** 279–306 (1994).

213. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58,** 267–288 (1996).

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

215. Van der Pas, S., Szabó, B. & van der Vaart, A. Adaptive posterior contraction rates for the horseshoe. *Electronic Journal of Statistics* **11,** 3196–3225 (2017).

216. Varian, H. R. Big data: New tricks for econometrics. *Journal of Economic Perspectives* **28,** 3–28 (2014).

217. Vavra, J. Inflation dynamics and time-varying volatility: New evidence and an ss interpretation. *Quarterly Journal of Economics* **129,** 215–258 (2013).

218. Wang, H. & West, M. Bayesian analysis of matrix normal graphical models. *Biometrika* **96,** 821–834 (2009).

219. Watson, M. W. Comment on ”The Empirical (Ir)Relevance of the Zero Lower Bound Constraint”. *NBER Macroeconomics Annual* **34** (2019).

220. West, M. Bayesian dynamic modelling. *Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian FM Smith,* 145–166 (2013).

221. West, M. Bayesian forecasting of multivariate time series: scalability, structure uncertainty and decisions. *Annals of the Institute of Statistical Mathematics* **72,** 1–31 (2020).

222. Wieland, J. F. & Yang, M.-J. Financial dampening. *Journal of Money, Credit and Banking* (2015).

223. Wolf, C. K. Masquerading Shocks in Sign-Restricted VARs. *Working Paper* (2016).

224. Wolf, C. K. Svar (mis-) identification and the real effects of monetary policy. *Working Paper* (2018).

225. Wooldridge, J. M. Score diagnostics for linear models estimated by two stage least squares. *Advances in econometrics and quantitative economics: Essays in honor of Professor CR Rao,* 66–87 (1995).

226. Yu, Y. & Meng, X.-L. To center or not to center: That is not the question—an Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics* **20,** 531–570 (2011).

227. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67,** 301–320 (2005).