From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2022

Essays on Estimating Impulse Responses in Macroeconometrics

Li, Dake

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp013r074z17w

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Essays on Estimating Impulse Responses in Macroeconometrics

Dake Li

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of Economics
Adviser: Christopher A. Sims

September 2022

© Copyright by Dake Li, 2022.

All Rights Reserved

Abstract

I develop and apply econometric methods for estimating impulse responses of macroeconomic variables to structural shocks.

Chapter 1 uses a structural vector autoregressive model with a Zero Lower Bound to characterize the censored nominal interest rate and the effect of unconventional monetary policy. Under empirically relevant non-Gaussian shocks, this chapter proposes a semiparametric identification scheme to prove point identification, without relying on the parametric form of the shock distribution. The key problem of model identification that this chapter solves is to deal semiparametrically with the non-linearity arising from censoring of the nominal interest rate at zero. An efficient Bayesian inference routine is designed to facilitate model estimation in practice. The empirical results suggest that the unconventional monetary policy has a small and transitory effect.

Chapter 2 is a joint work with Mikkel Plagborg-Møller and Christian K. Wolf, where we conduct a simulation study of Local Projection (LP) and Vector Autoregression (VAR) estimators of structural impulse responses across thousands of data generating processes, designed to mimic the properties of the universe of U.S. macroeconomic data. Our analysis considers various identification schemes and several variants of LP and VAR estimators. A clear bias-variance trade-off emerges: LP estimators have lower bias than VAR estimators but substantially higher variance at intermediate and long horizons. Consequently, unless researchers are overwhelmingly concerned with bias, shrinkage via Bayesian VARs or penalized LPs is attractive.

Chapter 3 is a joint work with Christopher A. Sims, where we estimate a structural VAR on a balanced panel of 10 countries to examine different causal channels that lead to comovement between aggregate credit and GDP. Among all the six structural shocks in our model, we find that there is one that pushes aggregate credit up and then later depresses GDP growth, but its effects are small in most countries. All the other five structural shocks still generate the positive or zero comovement. The shocks in our model are identified via variation across countries in their relative size, with their global trend taken into account. We also model the dependence between initial conditions and country constants to avoid the small-T panel bias.

Acknowledgements

I would like to take this chance to express my gratitude towards all the people who helped me in the process of my doctoral study.

First, I cannot exaggerate enough the tremendous help I received from my dissertation committee members, including Christopher A. Sims, Mikkel Plagborg-Møller, and Mark W. Watson. During the first two years of my study, it was their excellent teaching that made me deeply interested in macroeconometrics. Moreover, when I later started doing my research and exploring my career, they were always supporting me with insightful guidance. Throughout these six years, I gradually learned from them what a top researcher should look like.

Second, I also want to thank the entire Economics Department at Princeton. Through all kinds of seminars and meetings, the professors here were feeding me extremely diverse ideas on how to think about economic problems. Besides, the incredible assistance from the department staff was also crucial for every step of mine in this program. Most importantly, without the generous financial support from Princeton, my study here would not even be possible.

Last but not least, I sincerely appreciate the longstanding support from all my beloved families, especially my wife and my parents. Doing a Ph.D. was by no means easy, but I was lucky to have their trust and encouragement along this hard journey.

To my six-year life at Princeton.

Contents

1 Semiparametric Identification of SVAR Models with Zero Lower Bound **9**

1.1 Introduction . . . 9
1.2 Bivariate Model Setup . . . 14
1.2.1 Motivation for Censoring and Kink . . . 14
1.2.2 Specify the SVAR-ZLB Model with Two Regimes . . . 14
1.3 Lack of Point Identification under Gaussian Shocks . . . 17
1.3.1 Likelihood Evaluation for Any Shock Distribution . . . 17
1.3.2 Rotation of Likelihood Function for Gaussian Shocks . . . 18
1.3.3 Advantage of Specifying non-Gaussian shocks . . . 20
1.4 Point Identification under non-Gaussian Shocks . . . 22
1.4.1 Apply Independent Component Analysis through Hessian Matrices . . . 22
1.4.2 Reformulate as the Heckman Selection Model . . . 23
1.4.3 Use the Link between Structural Form and Reduced Form . . . 25
1.4.4 Takeaway from the semiparametric Identification Scheme . . . 26
1.5 Generalize to Multivariate SVAR-ZLB Models . . . 26
1.6 Bayesian Inference Routine and Simulation Study . . . 27
1.6.1 Bayesian Inference with a Gibbs Sampler . . . 28
1.6.2 Simulation Study . . . 29
1.7 Empirical Results . . . 30
1.8 Conclusion . . . 37
1.A Coherency Conditions on DGP . . . 38
1.B Proofs . . . 38
1.B.1 Proofs of Lemmas . . . 38
1.B.2 Proof of Theorems . . . 41
1.C Gibbs Sampling Scheme . . . 43
1.C.1 Symmetry in Mixture of Normal Distributions . . . 43
1.C.2 Prior . . . 43
1.C.3 Gibbs Sampler . . . 44
References . . . 46

2 Local Projections vs. VARs: Lessons From Thousands of DGPs **48**
2.1 Introduction . . . 48
2.2 The bias-variance trade-off . . . 51
2.2.1 Illustrative example . . . 51
2.2.2 Outlook . . . 55
2.3 Data generating processes . . . 55
2.3.1 Encompassing model . . . 55
2.3.2 DGPs and impulse response estimands . . . 56
2.3.3 Implementation . . . 57
2.3.4 Summary statistics . . . 59
2.4 Estimation methods . . . 61
2.5 Results . . . 63
2.5.1 There is a clear bias-variance trade-off between LP and VAR . . . 63
2.5.2 Shrinkage dramatically lowers variance, at some cost of bias . . . 65
2.5.3 No method dominates, but shrinkage is generally welcome . . . 66
2.5.4 SVAR-IV is heavily biased, but has relatively low dispersion . . . 69
2.5.5 Robustness . . . 69
2.5.6 Discussion: can we select the estimator based on the data? . . . 72
2.6 Conclusion and directions for future research . . . 73
2.A Details on DGP definitions . . . 75
2.A.1 Shock definition . . . 75
2.A.2 IV process calibration . . . 75
2.A.3 More persistent factor model . . . 75
2.B Details on estimation procedures . . . 76
References . . . 78

3 International Evidence on Credit and Economic Activity **81**
3.1 Introduction . . . 81
3.2 Data and model . . . 84
3.2.1 Data . . . 84
3.2.2 Model specification . . . 85
3.2.3 Identification . . . 87
3.3 Estimation procedure . . . 88
3.3.1 Dummy observations . . . 88
3.3.2 Other priors on parameters . . . 89
3.3.3 MCMC algorithm . . . 89
3.4 Main results . . . 90
3.4.1 Comovement of credit and GDP . . . 90
3.4.2 Size of global shocks and country shocks . . . 91
3.4.3 Comparison with and without global shocks . . . 94
3.5 Robustness checks . . . 95
3.5.1 Alternative sample . . . 95

3.5.2 Conventional identification scheme . . . 98
3.6 Conclusion and directions for future research . . . 98
3.A Details on MCMC algorithm . . . 99
3.B Shock size variation in global and country components . . . 100
3.C Cross correlation in monetary policy shocks . . . 102
References . . . 103

Chapter 1

Semiparametric Identification of SVAR Models with Zero Lower Bound

1.1 Introduction

Low nominal interest rates have been viewed as an important feature of the US economy in recent decades and among them the federal funds rate has even become frequently constrained at zero after the Great Recession.[1] The new feature of close-to-zero nominal interest rates has raised a challenge for central banks, which can no longer easily implement conventional interest rate cuts and have needed to turn to untested unconventional monetary policy tools. The Zero Lower Bound (ZLB) has thus given rise to increased research interest in the field of macroeconomics and has resulted in a huge debate on the effect of unconventional monetary policies among researchers (Kuttner, 2018; Bernanke, 2020).

An alternative to imposing many assumptions of fully-specified dynamic stochastic general equilibrium (DSGE) models is to adapt the model-robust structural vector autoregression (SVAR) framework to the ZLB setting to model macroeconomic data in the ZLB periods. Although conventional SVAR models can only deal with unconstrained data, Mavroeidis (2021) first incorporated ZLB into SVAR models to formulate this SVAR-ZLB model, to characterize the censored nominal interest rate and the unconventional monetary policy. Relative to conventional SVAR models, the key novelty of the SVAR-ZLB model setup is the non-linearity arising from censoring of the nominal interest rate at zero from below. He also presented how to use the ZLB as an additional identification channel to point-identify the model. However, this point identification strategy assumes that the short-run effect of the unconventional monetary policy is zero, which might not be true in reality.[2] Furthermore, the assumption of Gaussian shocks is crucial in Mavroeidis (2021) identification strategy because of the model's non-linearity, which limits generalization to other shock distributions. It is still unclear whether the SVAR-ZLB model in general can be point-identified and how this can be done without relying on the exact parametric form of the shock distribution.

[1]Author: Dake Li dakel@princeton.edu. I received helpful comments from Chris Sims, Mikkel Plagborg-Møller, Mark Watson, Bo Honoré, Ulrich Müller and Michal Kolesár. I am responsible for the remaining errors.

[2]For the sake of generality and convenience, in the rest of this paper, the SVAR-ZLB model by default refers to the model setup without zero-restrictions on the short-run effect of the unconventional monetary policy.

This paper proposes a generic semiparametric identification scheme for the SVAR-ZLB model under independent non-Gaussian shocks, providing point identification without relying on the parametric form of the shock distribution. First, this paper clearly explains the known result of no point identification with Gaussian shocks through the new lens of the likelihood function, thus demonstrating why there is an advantage in specifying non-Gaussian shocks while maintaining independence across shocks. Second, this paper develops a semiparametric identification scheme to point-identify the SVAR-ZLB model under non-Gaussian shocks, with the exact shock distribution unknown. To identify the structural parameters, this work adjusts the technique from the existing literature on independent component analysis (ICA), because the usual ICA technique only deals with the linear case and cannot handle the non-linearity arising from censoring. Moreover, this paper also designs an efficient Bayesian inference routine with a Gibbs sampler to facilitate model estimation in practice. Following this inference routine, the empirical analysis shows that the effect of the unconventional monetary policy is relatively small and transitory.

Censoring and kink in simultaneous equations, as two important features of this SVAR-ZLB model, make model identification challenging, compared to conventional SVAR models. Although Mavroeidis (2021) figured out an identification strategy when shocks are Gaussian, his point identification argument relies on the assumption of zero short-run effect of the unconventional monetary policy, and is also heavily connected to the parametric form of the Gaussian distribution, as in parametric sample selection models. When relaxing this assumption on the unconventional monetary policy, Mavroeidis claimed the lack of point identification by counting the number of parameters and the first and second moments. In this paper, I set up the SVAR-ZLB model with endogenously switched regimes as in Aruoba et al. (2021) and revisit the Gaussian setup from the perspective of likelihood. After a generic likelihood evaluation formula is provided, the known result of no point identification is interpreted through the lens of the joint Gaussian likelihood. Furthermore, this work utilizes the circular symmetry of the contour in the shock space to conclude that a rotation of the true SVAR-ZLB model will fit the data equally well. However, this is only a special case for Gaussian shocks, and specifying non-Gaussian shocks will break the circular contour and give a potential to point-identify the model.

The main contribution of this paper is a generic semiparametric point identification scheme for the SVAR-ZLB model in the case of independent non-Gaussian shocks, which does not need to rely on the exact parametric form of the shock distribution. This paper thus proves that, under some mild regularity conditions, the model can be point-identified once the Gaussian shocks are ruled out. The whole semiparametric identification scheme can be decomposed into three steps. First, this paper uses the non-Gaussianity in shocks and applies the ICA technique to identify the impact matrix, namely the structural parameters that represent the short-run effects of shocks. To handle the truncated-support problem arising from censoring in the model, I extend the conventional ICA technique by exploiting the structure of the Hessian matrix of the log density of the data, conditional on being away from the ZLB. Second, this paper reformulates the reduced form of the SVAR-ZLB model into a three-equation semiparametric Heckman selection model (Heckit) to identify the reduced-form coefficients for the lagged variables in the context of censoring. Third, this paper also points out the special link between the structural and reduced form of this SVAR-ZLB model across two endogenously switched regimes, which is crucial in identifying the kink in the model, namely the effect of the unconventional monetary policy.

Within this semiparametric identification scheme, the big challenge is how to apply ICA in the context of

censoring. Although the SVAR-ICA literature has shown how to use independent non-Gaussian shocks for model identification (Lanne et al., 2017; Gourieroux et al., 2017), researchers have mostly used the Darmois-Skitovich theorem, which can only deal with the unconstrained linear case. However, for the SVAR-ZLB model, censoring at ZLB brings in non-linearity and gives a truncated support in the shock space, which makes the shocks dependent conditional on being uncensored. This poses a big challenge, because the theoretical foundation of the SVAR-ICA literature, namely the Darmois-Skitovich theorem, breaks down in this context of censoring. Nevertheless, building on the idea of Lin (1998), this paper examines ICA through Hessian matrices of log densities and adjusts the ICA technique to deal with the truncated support in the shock space. With the aid of Hessian matrices, ICA is discussed from the likelihood perspective to show that censoring does not hinder identification through non-Gaussianity. Furthermore, by examining ICA through Hessian matrices, this paper mathematically links the ICA technique to the identification-through-heteroskedasticity technique, which is a commonly used identification strategy in the conventional SVAR literature.

This semiparametric identification scheme suggests three different strategies for estimating the SVAR-ZLB model in practice. First, people can use this semiparametric identification scheme directly to estimate the SVAR-ZLB model. Without relying on the parametric form of the shock distribution, robust semiparametric estimators can be derived from this semiparametric identification scheme. Nevertheless, there is a cost for the unknown parametric form because the semiparametric estimator might not be efficient for a moderate sample size in practice and will require researchers to choose among several possible semiparametric estimators and associated tuning parameters. Second, if researchers know the parametric form of the shock distribution, they can simply implement maximum likelihood estimation, in which the point identification will be automatically guaranteed by this semiparametric identification scheme, for any non-Gaussian shock distribution under mild regularity conditions. Finally, researchers can also use a flexible parametric distribution, such as a mixture of normal distributions, to approximate the unknown shock distribution and then implement maximum likelihood estimation in this parametric setting.

To facilitate model estimation in practice, this paper designs an efficient Bayesian inference routine for researchers to use. First, the unknown shock distribution is approximated with a mixture of normal distributions, following the spirit of the Sieve approach. Then, the data augmentation technique and the conjugate priors in my Bayesian inference framework are used to propose a Gibbs sampler, which can make posterior draws on the parameters and the augmented data in an alternating way, and compute the Bayesian posterior densities efficiently. Moreover, this Bayesian inference routine allows researchers to take advantage of standard Bayesian tools such as imposing shrinkage on the autoregressive coefficients and expressing prior beliefs about sign or magnitude constraints on impulse responses. In a simulation study with a calibrated trivariate model with non-Gaussian shocks, this paper finds that, even under a relatively dispersed prior, this model can be precisely estimated if it is identified through non-Gaussianity, especially when the sample size is relatively large and the occurrence of ZLB is not rare.

This paper also runs an empirical analysis on three key US economic variables, using the efficient inference routine. By fitting the SVAR-ZLB model to the time series of the federal funds rate, the output growth and the inflation rate, the results demonstrate a big and persistent effect of conventional monetary policy but a small and transitory effect of the unconventional monetary policy, given the same magnitude of a monetary policy shock. For the estimated shock distribution, it tends to have a severe fat tail, which supports the

non-Gaussianity assumption and helps with model identification. Furthermore, the shadow interest rate, which represents the desired nominal interest rate that central banks would like to select when ignoring the ZLB, is also imputed through the model and reveals a small magnitude across the ZLB periods.

Literature Review. There has been a long-standing debate about the effectiveness of unconventional monetary policies in the macroeconomics literature (Kuttner, 2018; Bernanke, 2020). Eggertsson & Woodford (2003) argued that quantitative easing would be ineffective if the expectation about the future enforcement of policies is not changed. However, Debortoli et al. (2020) provided evidence that ZLB is irrelevant to the economy using their empirical study and their theoretical models, as long as the unconventional monetary policy is well tuned to follow a shadow rate rule. Sims & Wu (2019) also concluded in their four-equation New Keynesian model that engaging in quantitative easing significantly mitigates the costs of a binding ZLB. In addition, Gertler & Karadi (2011) determined that the unconventional monetary policy can offset disruptions in private financial intermediaries and bring net benefits when the ZLB binds.

To flexibly model the ZLB data, this paper builds on Mavroeidis (2021) and Aruoba et al. (2021) to set up a SVAR-ZLB model that has two endogenously switched regimes. The ZLB constraint for nominal interest rates has inspired the econometrics literature to combine SVAR models with ZLB. One obvious weakness of conventional SVAR models is that variables are unconstrained, and the zero nominal interest rate is treated as a regular unconstrained data point, which will lead to estimation bias as in standard Tobit models (Amemiya, 1984). Mavroeidis (2021) incorporated ZLB into SVAR models to characterize the censoring of the nominal interest rate and the different effects of conventional and unconventional monetary policies. In the SVAR-ZLB model, the zero nominal interest rate in fact implies a negative unobserved shadow interest rate, which represents the desired stance of central banks if there is no ZLB. Furthermore, the effect of the negative shadow interest rate is allowed to be different from that of the positive nominal interest rate, to show the distinction between unconventional and conventional monetary policy. Aruoba et al. (2021) derived this econometric tool from approximating New Keynesian models and further generalized the setup. Aruoba et al. (2021) also formalized two endogenously switched regimes in this SVAR-ZLB model, to characterize the different parameters for normal and ZLB periods. The regime switching is endogenous because the regime is directly defined through the nominal interest rate.

In contrast to the SVAR-ZLB literature, the novelty of this paper is its presentation of how to identify the SVAR-ZLB model without the assumption of the unconventional monetary policy and the parametric form of the shock distribution. Mavroeidis (2021) used the ZLB as an additional identification channel under Gaussian shocks to point-identify all the structural parameters, based on the assumption that the unconventional monetary policy has zero short-run effect. When relaxing this assumption on unconventional monetary policy, he claimed the lack of point identification, because the number of parameters he had is more than the number of the first and the second moments he could use.[3] It is still unclear how this lack of point identification can be understood from the likelihood perspective and how to link it to the identification problem in conventional SVAR models. More importantly, whether we can identify this model without restrictions on the unconventional monetary policy, even without knowing the exact shock distribution, is still a question.

[3]Mavroeidis (2021) referred to this lack of point identification as set identification, but we need to be aware that the set is identified solely from the coherency conditions of the DGP and the identified set is usually quite wide.

The key technical challenge this paper solves is how to apply ICA in the context of censoring. Comon (1994) formalized the idea of ICA to untangle the independent components in a linear combination. Lanne et al. (2017) first introduced ICA to linear SVAR models to rigorously identify the structural parameters. To understand model identification through non-Gaussian shocks, Hyvarinen et al. (2001), Sims (2020) and Jarocinski (2021) depicted the star-shaped contour of independent non-Gaussian distributions and explained the intuition. Furthermore, Gourieroux et al. (2017) and Fiorentini & Sentana (2020) studied pseudo maximum likelihood estimators when the non-Gaussian shock distributions in SVAR models are misspecified. The linear SVAR-ICA approach has recently been used in several applied papers, with applications to the oil market (Braun, 2021) and the effects of monetary policies (Jarocinski, 2021). However, most of this SVAR-ICA literature (Lanne et al., 2017; Gourieroux et al., 2017) used the Darmois-Skitovich theorem to prove identification, which can only work for the unconstrained linear case and cannot handle the non-linearity from censoring. In particular, censoring generates a truncated joint support of independent shocks, and the shocks are no longer independent conditional on being uncensored. Unfortunately, the non-linearity from censoring cannot even be solved using the techniques in the nonlinear ICA literature (Gunsilius & Schennach, 2021), because there is no one-to-one mapping from the censored observations to the shocks. Nevertheless, Lin (1998) provided another perspective to examine linear ICA, i.e., applying ICA through Hessian matrices of log densities, which can be adjusted to work for the truncated joint support in this paper.

Because of the model's non-linearity from censoring, the proposed semiparametric identification scheme also exploits the semiparametric estimation of Heckman selection models, to handle the ZLB of the nominal interest rate and the regime-based selection of data. Amemiya (1984) described how to specify different sample selection models, including the three-equation Heckman selection model used in this paper. Once the sample selection model is specified, a semiparametric estimation method is needed. Powell (1984) gave a robust estimator, namely the censored least absolute deviation (CLAD) estimator, for the censored regression, which is used in this work to estimate the selection equation. Honore et al. (1997) developed a symmetric trimming technique to handle Heckman selection models when error terms are assumed to be symmetrically distributed. In Newey et al. (1990), they summarized different semiparametric estimation procedures using nonparametric regression techniques for Heckman selection models, including estimation based on kernel estimators (Robinson, 1988) and on series approximation (Cosslett, 1984), to estimate the outcome equation. Chamberlain (1986) further gave the asymptotic efficiency bound on the semiparametric estimation of Heckman selection models under different conditions.

Outline. The rest of this paper is organized as follows. Section 1.2 describes a bivariate setup of the SVAR-ZLB model. Section 1.3 interprets the lack of point identification under Gaussian shocks from the likelihood perspective. Section 1.4 proposes the semiparametric identification scheme for non-Gaussian shocks. Section 1.5 generalizes the identification argument to a multivariate setting. Section 1.6 discusses the efficient Bayesian inference routine in practice and gives a simulation study on the model estimation. Section 1.7 demonstrates the empirical results using the efficient inference routine.

1.2 Bivariate Model Setup

This section presents how to set up a SVAR-ZLB model to characterize ZLB of the nominal interest rate and the effect of unconventional monetary policy. A simple bivariate setting is used first in this section and Section 1.3 - Section 1.4 to simplify the illustration.[4] Section 1.2.1 motivates the two key features of a SVAR-ZLB model. Section 1.2.2 specifies the model in detail in the form of two different regimes.

1.2.1 Motivation for Censoring and Kink

I first use one motivation in the SVAR-ZLB literature to show the two key features we need to consider when involving ZLB into SVAR models. A simple linearized macroeconomic structure indicates that ZLB will generate the censoring and the kink in simultaneous equations, which are the two key features to be modeled in the SVAR-ZLB setup.

Mavroeidis (2021) offers a good motivation for how to model SVAR with ZLB, which is briefly illustrated here. If we consider the nominal interest rate r_t, the shadow interest rate r_t^*, the inflation rate π_t and log-deviation of the long-term bond quantity $b_{L,t}$. The shadow interest rate represents the desired nominal interest rate that central banks want to select when ignoring ZLB. A simple linearized macroeconomic structure has the following form

$$r_t^* = c_1 + \gamma_1 \pi_t + \sigma_1 \varepsilon_{1t} \tag{1.1}$$

$$\pi_t = c_2 + \gamma_2 r_t + \phi b_{L,t} + \sigma_2 \varepsilon_{2t} \tag{1.2}$$

$$b_{L,t} = \min\{\alpha r_t^*, 0\} \tag{1.3}$$

$$r_t = \max\{r_t^*, 0\} \tag{1.4}$$

where (1.1) - (1.4) represent the Taylor rule, the private sector equation, the quantitative easing and ZLB respectively.

After we plug (1.3) into (1.2), the private-sector equation will become

$$\pi_t = c_2 + \gamma_2 r_t + \gamma_2^* \min\{r_t^*, 0\} + \sigma_2 \varepsilon_{2t} \tag{1.5}$$

where we define $\gamma_2^* = \phi\alpha$. Note that there is a kink between γ_2 and γ_2^* in (1.5), which represent the two different effects of the interest rate when above or below zero, namely the effect of conventional monetary policy and the effect of unconventional monetary policy. Therefore, (1.1), (1.4) and (1.5) constitute a set of simultaneous equations with the censoring and the kink, to describe the macroeconomic variables that may potentially subject to the ZLB constraint.

1.2.2 Specify the SVAR-ZLB Model with Two Regimes

Now I formalize the setup of the SVAR-ZLB model which has the two above-mentioned features and specify the model in terms of two regimes to facilitate our discussion. In contrary to the exogenous regime switching

[4] A generalized multivariate setting is discussed in Section 1.5.

in the SVAR literature (Sims & Zha, 2006), the two regimes in the SVAR-ZLB will switch endogenously based on the level of the nominal interest rate.

We first consider a simple bivariate SVAR-ZLB model and use the conventional notations in SVAR models, for the observables $y_t = (y_{1t}, y_{2t})'$, where y_{1t} is nominal interest rate and y_{2t} is a private-sector variable (e.g. the inflation rate). In addition, y_{1t}^* denotes the shadow interest rate which is the latent interest rate not being censored. To facilitate our specification on the SVAR-ZLB model, I follow the regime-switching idea in Aruoba et al. (2021) to define two regimes using the regime indicator

$$s_t = \mathbf{1}\{y_{1t} > 0\} \tag{1.6}$$

where $s_t = 1$ means the standard regime (i.e. normal periods) and $s_t = 0$ means the ZLB regime (i.e. ZLB periods).

The data in two regimes are modeled differently. When in the standard regime (i.e. $y_{1t} > 0, s_t = 1$), we model the data as

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} y_{1t}^* \\ y_{2t} \end{bmatrix} = Bx_t + \varepsilon_t \tag{1.7}$$

whereas in ZLB periods (i.e. $y_{1t} = 0, s_t = 0$) we model the data as

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21}^* & A_{22} \end{bmatrix} \begin{bmatrix} y_{1t}^* \\ y_{2t} \end{bmatrix} = Bx_t + \varepsilon_t \tag{1.8}$$

The censoring at ZLB delivers the constrained nominal interest rate[5]

$$y_{1t} = \max\{y_{1t}^*, 0\} \tag{1.9}$$

For simplicity, we can denote $y_t^* = (y_{1t}^*, y_{2t})$ and further abbreviate the two impact matrices in (1.7) and (1.8) as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A^* = \begin{bmatrix} A_{11} & A_{12} \\ A_{21}^* & A_{22} \end{bmatrix} \tag{1.10}$$

A few things need to be clarified for the model specification (1.7) to (1.9). First, the lagged terms we use in (1.7) and (1.8) are only the lagged observables, i.e. $x_t = (y_{t-1}', y_{t-2}', \cdots, y_{t-p}', 1)'$.[6] Second, the structural shocks $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ are exogenous, not influenced by the censoring, and stay independent across time and across coordinates with their respective density being f_i, i.e. $\varepsilon_{it} \overset{\text{iid}}{\sim} f_i, i = 1, 2$. Thrid, as in the SVAR literature, this model is subject to a normalization issue and here we normalize the diagonal of A and A^* to

[5] Even if y_{1t} is censored at $b \neq 0$, we can shift down the time series of y_{1t} by b and accordingly adjust the constant term in the true SVAR model. Thus, assuming censorship at 0 makes the modeling simple and without loss of generality. See appendix for the alternative modeling with $b \neq 0$.

[6] Including only the lagged observables and excluding the lagged shadow interest rate makes the model more tractable. Furthermore, Aruoba et al. (2021) showed that their New Keynesian model with ZLB is well approximated by a SVAR-ZLB model with only the lagged observables included on the right-hand side.

be 1, i.e. $A_{11} = A_{22} = 1$.

Two minor issues regarding simultaneous equation models with two regimes as in (1.7) and (1.8) are continuity and coherency of the data generating process (DGP). This paper restricts the model to satisfy the continuity condition and the coherency condition, which are two minimum conditions on the DGP and commonly used in the SVAR-ZLB literature (Mavroeidis, 2021; Aruoba et al., 2021), in order to guarantee the equilibrium in the model to be continuous and unique. The definition of continuity and coherency in the SVAR-ZLB model is as follows:

- Continuity. There should be no jump in y_{2t} across the two regimes when $y^*_{1t} = 0$.
- Coherency. There exists a unique regime being coherent at time t. Either $y_{1t} > 0$ or $y_{1t} = 0$ can be true.

The continuity condition is satisfied naturally in my model setup, because the only change in parameters across the two regimes is the kink between A_{21} and A^*_{21}. For the coherency condition, I simply need to restrict the domain of A and A^* and force the determinants of these two impact matrices to be of the same sign, i.e.

$$\text{sign}(|A|) = \text{sign}(|A^*|) \tag{1.11}$$

A detailed discussion on these two minimum conditions can be found in Section 1.A. With these two minimum conditions on the DGP, the model now will always yield one unique equilibrium, which is also continuous in terms of the shocks.

After specifying the model, we can now interpret the economic meanings in the SVAR-ZLB model. In (1.7) and (1.8), the first row represents the monetary policy equation and the second row represents the private-sector equation. For the parameters, the matrix for the lagged terms B remains the same across regimes, and the only change among the parameters across regimes is the kink between A_{21} and A^*_{21}, which captures the change in the short-run effect of monetary policy from the conventional one A_{21} to the unconventional one A^*_{21}. The long-run effects of monetary policy will also be different when we feed different short-run effects into the model. For the variables, both the lagged terms x_t and the exogenous shocks ε_t are unchanged across regimes, and the only change happens in the kink of $(y^*_{1t}, y_{2t})'$.

To facilitate our discussion of the model identification later in this paper, I also use the regime-contingent notations as in Aruoba et al. (2021) to rewrite the model. We can combine the two regimes in (1.7) and (1.8) into one regime-contingent model

$$A(s_t)y_t(s_t) = Bx_t + \varepsilon_t \tag{1.12}$$

$$y^*_t = y_t(s_t) \tag{1.13}$$

where the regime-contingent impact matrix is

$$A(s_t) = s_t \cdot A + (1 - s_t) \cdot A^* \tag{1.14}$$

and $y_t(s) = (y_{1t}(s), y_{2t}(s))'$ is defined as the latent outcome generated using the parameters in regime s. Note that the DGP will choose y^*_t to be one of the two sets of latent outcomes $y_t(1)$ and $y_t(0)$, and which

latent outcome gets chosen will recursively depend on the sign of the chosen y_{1t}^* through (1.9). Thus, in contrast to the regime-switching SVAR literature (Sims & Zha, 2006), the regime s_t here is endogenously switching based on the endogenous variable y_{1t}.[7]

After building the regime-contingent structural-form equations in (1.12), we can also derive the regime-contingent reduced-form equations,

$$y_t(s_t) = A(s_t)^{-1}Bx_t + A(s_t)^{-1}\varepsilon_t = \beta(s_t)x_t + u_t(s_t) \tag{1.15}$$

where $\beta(s_t) = A(s_t)^{-1}B$ are the reduced-form lag coefficients and intercepts, $u_t(s_t) = G(s_t)\varepsilon_t$ are the reduced-form errors and $G(s_t) = A(s_t)^{-1}$ translating the structural shocks to the reduced-form errors. With the similar notation rule as in (1.14), I denote $\beta = \beta(1)$, $\beta^* = \beta(0)$, $G = G(1)$ and $G^* = G(0)$ to simplify the illustration of the proof later.

1.3 Lack of Point Identification under Gaussian Shocks

This section interprets why we cannot achieve point identification of the SVAR-ZLB model under Gaussian shocks, from the likelihood perspective. Section 1.3.1 first provides the generic formulas to evaluate the likelihood of the SVAR-ZLB model for any shock distribution. Section 1.3.2 points out that the key reason of no point identification in the Gaussian case is the circular contour which makes the likelihood unchanged after a rotation of the SVAR-ZLB model. Section 1.3.3 illustrates the advantage of specifying more realistic non-Gaussian shocks for the purpose of point identification.

1.3.1 Likelihood Evaluation for Any Shock Distribution

For any shock distribution, there are two ways to generically evaluate the likelihood of the SVAR-ZLB model. One way is to directly use observables and evaluate in the y-space, and another way is to transform observables to shocks and evaluate in the ε-space. The first way is useful to compute the likelihood in practice, whereas the second way is important to understand the identification problem with Gaussian or non-Gaussian shocks.

We first discuss how to evaluate the likelihood directly in the y-space. For simplicity, we only focus how to compute the likelihood of $y_t|x_t$ in this subsection. We denote the joint distribution of the two shocks as $\mathbf{f}(\varepsilon_t) = f_1(\varepsilon_{1t}) \cdot f_2(\varepsilon_{2t})$. One the one hand, the uncensored observation y_t in the standard regime $s_t = 1$ gives

$$p(y_t, s_t = 1|x_t) = |A| \cdot \mathbf{f}(Ay_t - Bx_t) \tag{1.16}$$

On the other hand, the censored observation $y_t = (0, y_{2t})'$ in the ZLB regime $s_t = 0$ implies a negative

[7] With the coherency condition, $y_{1t}(1) > 0$ implies $y_{1t}(0) > 0$ and vice versa. Thus, there exists one unique regime that can be coherently chosen in the DGP.

shadow interest rate y_{1t}^* and involves integral in the tail,

$$p(y_t, s_t = 0|x_t) = \int_{-\infty}^{0} |A^*| \cdot \mathbf{f}(A^*(y_{1t}^*, y_{2t})' - Bx_t) dy_{1t}^* \tag{1.17}$$

Therefore, the full likelihood of $y_t|x_t$ is comprised of (1.16) and (1.17),

$$p(y_t|x_t) = s_t \cdot p(y_t, s_t = 1|x_t) + (1 - s_t) \cdot p(y_t, s_t = 0|x_t) \tag{1.18}$$

Besides the evaluation in the y-space, we can also evaluate the likelihood in the ε-space. For the standard regime $s_t = 1$, we can easily impute the shocks $\varepsilon_t = Ay_t - Bx_t$ and plug it into the shock distribution $\mathbf{f}(\varepsilon_t)$. For the ZLB regime, we will need to impute the pseudo shocks $\varepsilon_t^0 = A^*(0, y_{2t})' - Bx_t$, and integrate out the true shocks subject to a linear constraint. The details of likelihood evaluation in the ε-space can be found in Lemma 1.B.1.

The likelihood evaluation formulas in the y-space and in the ε-space are both useful in this paper. Although these two formulas are equivalent as shown in Lemma 1.B.1, the evaluation in the y-space is easier to compute the likelihood in practice in Section 1.6 and Section 1.7, whereas the evaluation in the ε-space is more important for us to understand the identification problem with Gaussian or non-Gaussian shocks in Section 1.3.2 and Section 1.3.3.

1.3.2 Rotation of Likelihood Function for Gaussian Shocks

When we take the common assumption that the shocks are Gaussian, the SVAR-ZLB model will not be point-identified, because the circular contour of the Gaussian distribution enables us to preserve the likelihood value when we rotate this model, even in the case of censoring. A rotation of the SVAR-ZLB model means the rotation of everything in the ε-space when we evaluate the likelihood.

The existing literature (Mavroeidis, 2021) relied on an additional assumption to achieve the point identification of this model under Gaussian shocks, i.e. the short-run effect of unconventional monetary policy is exactly zero.[8] However, this assumption might not hold in reality and the identification strategy based on this assumption can be fragile. When we relax this assumption and want to estimate the short-run effect of unconventional monetary policy, the literature concluded the lack of point identification (a.k.a. under-identification) in this model, by counting the number of first and second moments and unknown parameters. This paper tries to interpret this under-identification problem from the new perspective of likelihood.

I first use two figures to revisit the likelihood evaluation in y-space and ε-space for the Gaussian shocks. First for the y-space, as shown in the left panel of Figure 1.1, (y_{1t}^*, y_{2t}) follows a joint Gaussian distribution whose contour is colored in red when $s_t = 1$, and follows another joint Gaussian distribution whose contour is colored in blue when $s_t = 0$. One uncensored observation for $s_t = 1$ is shown as a red point, and another censored observation for $s_t = 0$ is shown as a blue point. Due to the censoring, the observation for $s_t = 0$ always appears on the ZLB boundary line (colored in green). Moreover, the likelihood calculation for the censored observation involves integrating out the tail along the integral line (denoted as a blue arrow). Second

[8] For the model setup as in Section 1.2, this assumption will even imply the long-run effect of unconventional monetary policy is also zero.

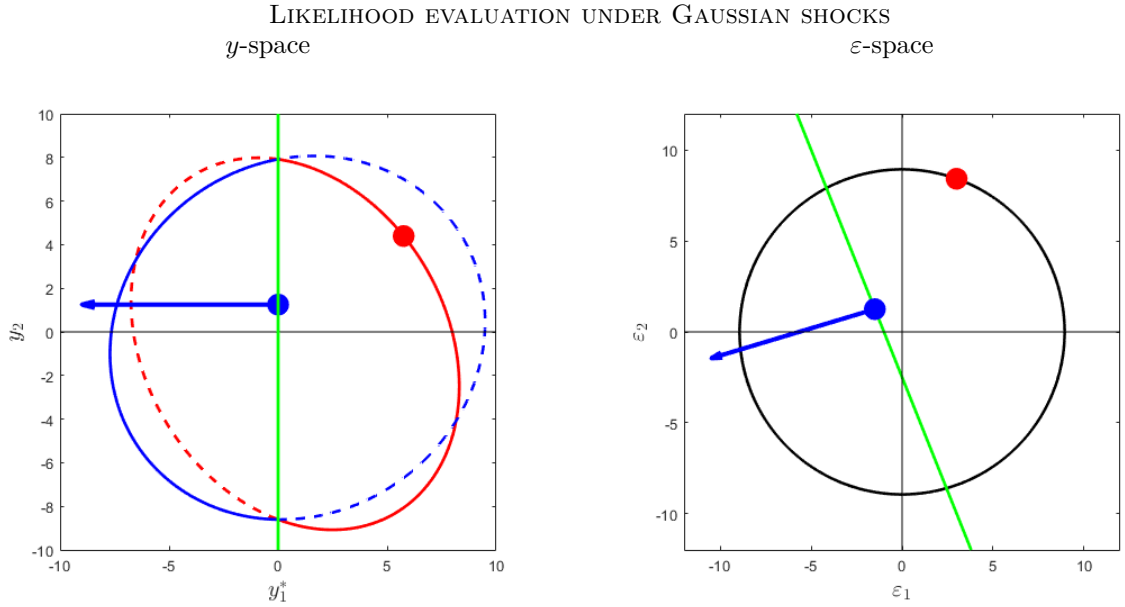


Figure 1.1: The observation is depicted as a red point for the standard regime and as a blue point for the ZLB regime. The joint distribution is depicted as a red ellipse for the standard regime and as a blue ellipse for the ZLB regime. The contour of the shocks, distributed as independent standard Gaussian, is plotted as a black circle. The green line is the ZLB boundary line, and the blue arrow is the integral line towards the tail.

we can translate the uncensored and the censored observation, the ZLB boundary line and the integral line into the ε-space, which are depicted in the right panel of Figure 1.1 using the same legend scheme. Note that the contour in the ε-space is circular, because of the independent standard Gaussian distribution of the shocks.[9] The green line represents the set of shocks that are marginally binding at ZLB and the blue arrow represent those possible true shocks that can match the observed y_{2t} and imply a negative shadow interest rate.[10]

The circular contour for the Gaussian distribution in the ε-space sheds light on why we cannot achieve point identification of the model under Gaussian shocks. If we are able to rotate everything in the right panel of Figure 1.1, including the uncensored and the censored observation, the ZLB boundary line and the integral line, by the same angle, we will actually change all the structural parameters of the SVAR-ZLB model while preserving the likelihood value due to the circular symmetry of the Gaussian distribution. From this likelihood perspective, we can link the under-identification problem of SVAR-ZLB models to the well-known under-identification problem of conventional SVAR models, if shocks are assumed to be Gaussian.[11] However, in order to make A and A^* only differ in the bottom-left corner, the challenge is that we cannot

[9] To simplify the discussion in this section, we temporarily change to a new normalization scheme where the variance of these Gaussian shocks are normalized to be 1, instead of forcing the diagonal of A and A^* to be 1. The details of the mapping between these two normalization scheme can be found in Lemma 1.B.2.

[10] For the detailed formulas of the ZLB boundary line and the integral line in the right panel of Figure 1.1, see the remark of Lemma 1.B.1.

[11] There is a small difference when we rotate the SVAR-ZLB model, that is some rotations will be ruled out because of violating the coherency condition (1.11). All the rotations that make the model satisfy the coherency condition yield the identified set of the structural parameters. We need to be cautious that the identified set can often be quite wide in this under-identification problem.

simply rotate A^* but we still need to rotate the integral line. Fortunately, I show in Lemma 1.B.3 that we can follow a specific mapping rule to transform A^* to both keep the structural connection with A and get the integral line rotated by the same angle.

Theorem 1.1. *Under Gaussian shocks, all the structural parameters (A, A^*_{21}, B) in the SVAR-ZLB model will not be point-identified.*

See proof in Section 1.B.

Theorem 1.1 formally proves that there is no way to achieve point identification of the SVAR-ZLB model simply because of the rotation in the circular contour.[12] In contrast with Mavroeidis (2021), instead of assuming $A^*_{21} = 0$ to achieve point identification, I remove this assumption in order to directly estimate the short-run effect of unconventional monetary policy and end up with no point identification. If the shocks are indeed Gaussian, the only way to point-identify this model is to make further assumptions. For example, one can follow Mavroeidis (2021) to make assumptions on unconventional monetary policy, or follow the SVAR literature to impose zero-restrictions on A, or even assume stochastic volatility of shocks.[13] Moreover, the lack of point identification can also appear when shocks follow any elliptical distribution with a diagonal dispersion matrix, where the contour is again circular and the same rotation logic can be applied. But it is noteworthy that the Gaussian distribution is the only elliptical distribution that generate independent shocks and have a diagonal dispersion (covariance) matrix.

1.3.3 Advantage of Specifying non-Gaussian shocks

If we get rid of the assumption that shocks are exactly Gaussian and take a more realistic stance on the non-Gaussian shocks, this will bring a big advantage on identifying the SVAR-ZLB model. The lack of point identification in Section 1.3.2 is a problem special to the case when shocks are exactly Gaussian, which is an unrealistic assumption compared to many findings in the applied research favoring non-Gaussian shocks. In addition, the SVAR-ICA literature provides solid foundation for using non-Gaussian shocks to identify conventional SVAR models and can be adjusted to help identify SVAR-ZLB models.

There are many findings in the SVAR literature that support the non-Gaussianity of shocks. Among them, Brunnermeier et al. (2021) used a regime-switching SVAR model to fit the US time series and found that the t-distribution fits the shocks better than the Gaussian distribution. The big shocks in their model tend to appear in one coordinate at a time, which makes the independent t-distribution a good specification for the shocks. Many recent empirical papers generate empirical findings with their shocks and their impulse responses directly identified trhough non-Gaussianity (Braun, 2021; Jarocinski, 2021).

The SVAR-ICA literature has also pointed out how to use non-Gaussian shocks to identify SVAR in the linear unconstrained case. For independent non-Gaussian shocks, Lanne et al. (2017) relies on the Darmois-Skitovich Theorem to identify the SVAR model. Gourieroux et al. (2017) further proved that this

[12] Mavroeidis (2021) proved the under-identification when relaxing the assumption of A^*_{21}, by counting the number of the available moments and the parameters, but it is unclear whether we have exhausted all the useful moments. This paper discusses the under-identification from the likelihood perspective and formally rules out the identification through any other potentially useful moments.

[13] Aruoba et al. (2021) did not use the assumption $A^*_{21} = 0$ but assumed stochastic volatility of shocks. Hence their model is identified through heteroskedasticity, instead of through the help of ZLB. Moreover, in Section 4.1 of their paper, they also tried a model under homoskedastic shocks, which I believe to be under-identified due to Theorem 1.1 of my paper.

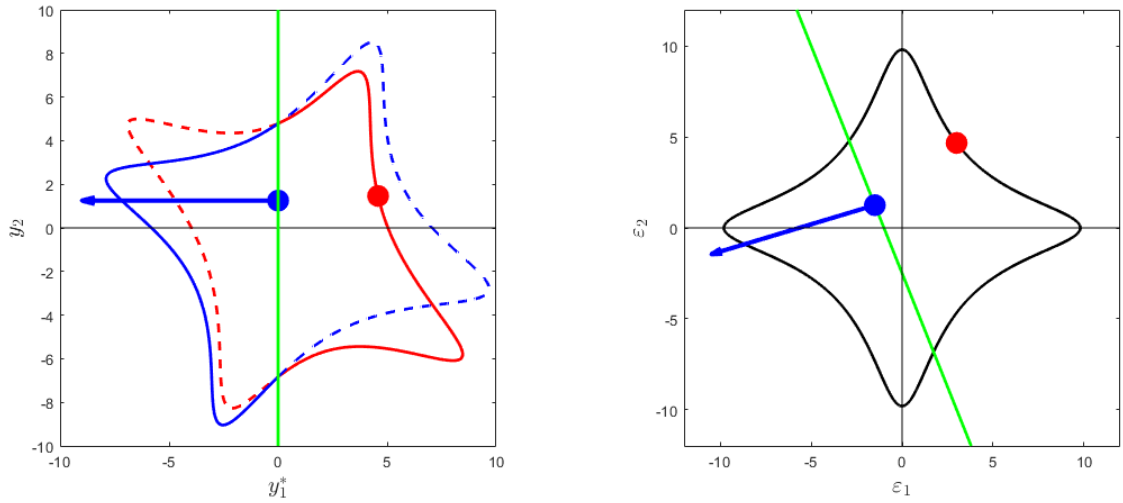


Figure 1.2: The observation is depicted as a red point for the standard regime and as a blue point for the ZLB regime. The joint distribution is depicted as a red ellipse for the standard regime and as a blue ellipse for the ZLB regime. The contour of the shocks, distributed as independent t_3 distribution, is plotted as a black curve. The green line is the ZLB boundary line, and the blue arrow is the integral line towards the tail.

identification strategy is fairly robust even when we slightly misspecify the shock distribution. Sims (2020) examined this identification strategy through the non-circular contour of the independent t-distribution, which shows a big advantage of specifying non-Gaussian shocks.

However, the SVAR-ICA literature does not know how to identify the model if we have non-linearity arising from the censoring at ZLB, and this paper tries to extend this literature to deal with the censoring problem. The key challenge is that the unconditionally independent shocks are no longer independent conditional on being not censored, due to the truncated support in the ε-space as in Figure 1.1, and thus the conventional wisdom in the SVAR-ICA literature cannot be directly applied. Following the spirit of Sims (2020), I use Figure 1.2, which is analog to Figure 1.1 in Section 1.3.2, to illustrate the likelihood of the SVAR-ZLB model under the independent t-distribution. If the shock is distributed as t-distribution with 3 degrees of freedom, the contour in the ε-space is now star-shaped and pointing out along each axis. Although we only have a truncated support on the right of the ZLB boundary line, that still intuitively seems to rule out the rotation problem in Section 1.3.2, since the star-shaped contour is not likely to preserve the likelihood after we rotate everything in the ε-space.[14] Nevertheless, we still need to have a rigorous identification scheme to prove that the truncated support with the non-circular contour indeed brings the identification of the SVAR-ZLB model for a generic non-Gaussian shock distribution, which is listed in Section 1.4.

[14]Note that the likelihood will remain the same if we rotate exactly by a multiple of 90 degrees, which is just changing the label and the sign of shocks.

1.4 Point Identification under non-Gaussian Shocks

This section proposes a semiparametric identification scheme for SVAR-ZLB models under non-Gaussian shocks, without relying on the parametric form of the shock distribution. The identification scheme is decomposed into three parts. First, Section 1.4.1 discusses how to adjust the ICA techniques through Hessian matrices of log densities to deal with the non-linearity problem arising from the censoring in the model. Second, Section 1.4.2 reformulates the SVAR-ZLB model as a three-equation Heckman selection model. Third, Section 1.4.3 uses the special link between the structural form and the reduced form of the model.

1.4.1 Apply Independent Component Analysis through Hessian Matrices

The first step is to identify the impact matrix A through non-Gaussian shocks. This paper adjusts the ICA argument through Hessian matrices and thus deal with the truncated support due to the censoring in the model. In addition, when we discuss ICA from the lens of Hessian matrices, the identification technique through non-Gaussian shocks becomes mathematically related to identification-through-heteroskedasticity.

Now we consider how to identify A when the shock distribution f_i is non-Gaussian with an unknown parametric form. The key challenge that prevents us from directly applying ICA is the truncated support in the ε-space, i.e. $\beta_1.x_t + G_1.\varepsilon_t \geq 0$, which makes $(\varepsilon_{1t}, \varepsilon_{2t})$ no longer independent. However, by borrowing ideas from Lin (1998), I achieve the identification of A using Hessian matrices of the log-likelihood function, at two interior points of this truncated support.

Theorem 1.2. *Suppose*

- *f_i has a full support on* $\mathbf{R}$
- $\log f_i \in C^2$
- *$(\log f_i)''$ is not constant (i.e. f_i is non-Gaussian)*

then A is point-identified using only the uncensored data given x_t, i.e. using $y_t|x_t, y_{1t} > 0$.

Proof. The log-likelihood of (y_{1t}, y_{2t}) conditional on x_t and $y_{1t} > 0$ can be written as,

$$\log p(y_t|x_t, y_{1t} > 0) = \log|A| + \log \mathbf{f}(\varepsilon_t) - \log \Pr(y_{1t} > 0|x_t) \tag{1.19}$$

where $p(y_t|x_t, y_{1t} > 0)$ is the conditional distribution in the interior of the truncated support, $\mathbf{f}(\varepsilon_t)$ is the unconditional shock distribution, $\Pr(y_{1t} > 0|x_t)$ is the probability of being uncensored.

Then we take second-order derivatives with respect to y_t on both sides of (1.19) and get

$$H_y(y_t) = A' H_\varepsilon(\varepsilon_t) A \tag{1.20}$$

where $H_y(y_t) = \frac{\partial^2 \log p(y_t|x_t, y_{1t}>0)}{\partial y_t \partial y_t'}$, and $H_\varepsilon(\varepsilon_t) = \frac{\partial^2 \log \mathbf{f}(\varepsilon_t)}{\partial \varepsilon_t \partial \varepsilon_t'}$.

The formerly independent shocks imply that $\log \mathbf{f}(\varepsilon_t) = \log f_1(\varepsilon_{1t}) + \log f_2(\varepsilon_{2t})$, and we have a diagonal matrix for $H_\varepsilon(\varepsilon_t)$:

$$H_\varepsilon(\varepsilon_t) = \begin{bmatrix} (\log f_1)''(\varepsilon_{1t}) & 0 \\ 0 & (\log f_2)''(\varepsilon_{2t}) \end{bmatrix} \tag{1.21}$$

Given the diagonal structure in $H_\varepsilon(\varepsilon_t)$, we can follow the technique similar to identification through heteroskedasticity. First, we pick two points in the interior of the truncated support, namely y_t and $\tilde{y}_t$ in the y-space, or equivalently ε and $\tilde{\varepsilon}_t$ in the ε-space. Then we compute the eigenvalue decomposition of the following term,

$$H_y(y_t)^{-1} H_y(\tilde{y}_t) = A^{-1} H_\varepsilon(\varepsilon_t)^{-1} H_\varepsilon(\tilde{\varepsilon}_t) A \tag{1.22}$$

where the left-hand side is the observed Hessian matrices, and the right-hand side is in the form of eigenvalue decomposition. The eigenvectors we collect from (1.22) gives the identification of A.

Note that we will not get a unique eigenvalue decomposition, if $H_\varepsilon(\varepsilon_t)$ is proportional to $H_\varepsilon(\tilde{\varepsilon}_t)$. However, if we get into this knife-edge situation, we can always get around it by slightly change one point we have picked. For example, we can keep the point $\tilde{\varepsilon}_t$ and then change the first coordinate of the point ε_t by $\Delta\varepsilon_{1t}$. Then we get the Hessian matrix at the new point $(\varepsilon_{1t} + \Delta\varepsilon_{1t}, \varepsilon_{2t})$ as

$$H_\varepsilon(\varepsilon_{1t} + \Delta\varepsilon_{1t}, \varepsilon_{2t}) = \begin{bmatrix} (\log f_1)''(\varepsilon_{1t} + \Delta\varepsilon_{1t}) & 0 \\ 0 & (\log f_2)''(\varepsilon_{2t}) \end{bmatrix} \tag{1.23}$$

which will no longer be proportional to $H_\varepsilon(\tilde{\varepsilon}_t)$. □

Theorem 1.2 proves the semiparametric identification of A using the property of the likelihood function in the model, but we do not rely on a specific parametric form of the shock distribution. Therefore, if the mild regularity conditions in Theorem 1.2 hold, the non-Gaussian shocks in the SVAR-ZLB model, which brings the non-circular contour with a truncated support, are enough to identify the impact matrix A in the SVAR-ZLB model. Since we only require two points in the truncated support to prove the identification of A, using all the points in the truncated support will even give over-identification of A.

1.4.2 Reformulate as the Heckman Selection Model

The next step is to identify the reduced-form parameters. To show how to estimate the reduced-form parameters β and β^* in the case of the censoring and the kink, I reformulate the SVAR-ZLB model as a three-equation Heckman selection model (Amemiya, 1984).

We can rewrite the regime-contingent SVAR model in the form of a three-equation Heckman selection model,

$$y_{1t}(1) = \beta_1 . x_t + u_{1t}(1) \tag{1.24}$$

$$y_{2t}(1) = \beta_2 . x_t + u_{2t}(1) \tag{1.25}$$

$$y_{2t}(0) = \beta_2^* . x_t + u_{2t}(0) \tag{1.26}$$

where $y_{1t}(1), y_{2t}(1), y_{2t}(0)$ are latent variables as in conventional Heckman selection models. In the selection equation (1.24), we only observe $y_{1t} = \max\{y_{1t}(1), 0\}$. In the observation equation (1.25), $y_{2t} = y_{2t}(1)$ is observed when $y_{1t}(1) > 0$. However, in contrary to conventional Heckman selection models, when $y_{1t}(1) < 0$, we can still observe $y_{2t} = y_{2t}(0)$ from another observation equation (1.26).

Theorem 1.3. *If the following regularity conditions*

- *shock distribution f_i is symmetric around 0*
- *both control functions $H_1(z) = E[u_{2t}(1)|u_{1t}(1) > z]$ and $H_0(z) = E[u_{2t}(0)|u_{1t}(1) \leq z]$ are not linear in z*
- *$E[x_t x_t']$ has full rank*

hold, then we can identify all the reduced-form parameters in the above-mentioned Heckman selection model (1.24) - (1.26), namely $\beta_{1\cdot}$, $\beta_{2\cdot}$ and $\beta^*_{2\cdot}$.

Proof. To identify the reduced-form parameters $\beta_{1\cdot}$, $\beta_{2\cdot}$ and $\beta^*_{2\cdot}$, we simply need to use the censored least absolute estimator (Powell, 1984) and the semiparametric estimator for Heckman selection models (Newey et al., 1990):

- Given (1.24), we rely on the symmetry distribution of $u_{1t}(1)$ and the censored least absolute estimator to identify $\beta_{1\cdot}$ through quantile regression,

$$\text{med}(y_{1t}|x_t) = \max\{\beta_{1\cdot} x_t, 0\} \tag{1.27}$$

- Given (1.24) and (1.25), we can identify $\beta_{2\cdot}$ through the semiparametric estimator for conventional Heckman selection models,

$$\begin{aligned} &E[y_{2t}|x_t, y_{1t} > 0] = E[y_{2t}(1)|x_t, y_{1t}(1) > 0] \\ =&\beta_{2\cdot} x_t + E[u_{2t}(1)|u_{1t}(1) > -\beta_{1\cdot} x_t] = \beta_{2\cdot} x_t + H_1(-\beta_{1\cdot} x_t) \end{aligned} \tag{1.28}$$

- Given (1.24) and (1.26), we can semiparametrically identify $\beta^*_{2\cdot}$ in the similar way,

$$\begin{aligned} &E[y_{2t}|x_t, y_{1t} = 0] = E[y_{2t}(0)|x_t, y_{1t}(1) \leq 0] \\ =&\beta^*_{2\cdot} x_t + E[u_{2t}(0)|u_{1t}(1) \leq -\beta_{1\cdot} x_t] = \beta^*_{2\cdot} x_t + H_0(-\beta_{1\cdot} x_t) \end{aligned} \tag{1.29}$$

□

Theorem 1.3 gives a routine for estimating the reduced-form parameters in the SVAR-ZLB model.[15] Because of the censoring, the OLS estimates of the reduced-form parameters will be biased, and an estimator involving the censoring and the sample selection is necessary to generate unbiased estimates. In this three-equation Heckman selection model, we simply run the censored absolute deviation estimator for the selection

[15] Note that $\beta^*_{1\cdot}$ does not appear in this three-equation Heckman selection model, and thus cannot be directly identified in this section. However, we can identify $\beta^*_{1\cdot}$ in Section 1.4.3.

equation (1.24), and run the semiparametric Heckit estimator twice, for the observation equation (1.25) and (1.26) separately. The estimators we use in this section again do not rely on the parametric form of the shock distribution.

1.4.3 Use the Link between Structural Form and Reduced Form

The last step of this identification scheme is to use the special link between the structural form and the reduced form in this SVAR-ZLB model to identify the remaining parameters, including B, $\beta_{1\cdot}^*$ and A_{21}^*.

After we semiparametrically identify A, $\beta_{1\cdot}$, $\beta_{2\cdot}$ and $\beta_{2\cdot}^*$ in Section 1.4.1 and Section 1.4.2, the remaining task is to identify B, $\beta_{1\cdot}^*$ and A_{21}^*. First, it is straightforward now to identify B, because we have $\beta = (\beta_{1\cdot}', \beta_{2\cdot}')'$ and $B = A\beta$. Second, for identifying $\beta_{1\cdot}^*$ and A_{21}^*, we can exploit the special link between the structural form and the reduced form of this SVAR-ZLB model. As mentioned in Section 1.2.2, although the reduced forms across the two regimes are largely different, the structural forms across the two regimes only differ between A_{21} and A_{21}^*, which leads to the identification of $\beta_{1\cdot}^*$ and the over-identification of A_{21}^*.

Theorem 1.4. *Given A, B, β and $\beta_{2\cdot}^*$, if*

- $A_{11} \neq 0$
- *there is at least one non-zero element in $\beta_{1\cdot}$*

then $\beta_{1\cdot}^*$ and A_{21}^* will be identified.

Proof. To identify $\beta_{1\cdot}^*$ and A_{21}^*, we rely on the special link between the structural form and the reduced form in the ZLB regime,

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21}^* & A_{22} \end{bmatrix} \begin{bmatrix} \beta_{1\cdot}^* \\ \beta_{2\cdot}^* \end{bmatrix} = \begin{bmatrix} B_{1\cdot} \\ B_{2\cdot} \end{bmatrix} \tag{1.30}$$

Notice that in (1.30), the only unknown parameters at this stage are $\beta_{1\cdot}^*$ and A_{21}^*. In the first place, we can directly identify $\beta_{1\cdot}^*$ from the first row of (1.30),

$$\beta_{1\cdot}^* = \frac{1}{A_{11}}(B_{1\cdot} - A_{12}\beta_{2\cdot}^*) \tag{1.31}$$

Then we can always find at least one non-zero element in $\beta_{1\cdot}^*$, if $\beta_{1\cdot}$ has at least one non-zero element, because they are proportional to each other as shown in (1.32)

$$\begin{aligned} \beta_{1\cdot}^* = G_{1\cdot}^* B = \frac{1}{|A^*|}(A_{22}, -A_{12}) \cdot B \\ = \frac{|A|}{|A^*|}\frac{1}{|A|}(A_{22}, -A_{12}) \cdot B = \frac{|A|}{|A^*|} G_{1\cdot} B = \frac{|A|}{|A^*|}\beta_{1\cdot} \end{aligned} \tag{1.32}$$

Now without loss of generality, suppose there is a non-zero element in the first column of $\beta_{1\cdot}^*$, i.e. $\beta_{11}^* \neq 0$. Then we can directly identify A_{21}^* from the second row of (1.30),

$$A_{21}^* = \frac{1}{\beta_{11}^*}(B_{21} - A_{22}\beta_{21}^*) \tag{1.33}$$

□

1.4.4 Takeaway from the semiparametric Identification Scheme

After proving the identification of the SVAR-ZLB model using this semiparametric identification scheme, I now discuss how it sheds light on the estimation of SVAR-ZLB models in practice. We will consider two different cases: when we know or when we do not know the parametric form of the non-Gaussian shock distribution.

If we do not know the parametric form of the non-Gaussian shock distribution, as we mentioned through Section 1.4.1 - Section 1.4.3, we can use this semiparametric identification scheme to directly estimate the SVAR-ZLB model. In this case, we will need to nonparametrically estimate the Hessian matrices of the log-density of y_t and then semiparametrically estimate the three-equation Heckman selection model. Both the bandwidth for estimators and the points to be selected for Hessian matrices will be a important practical choice that researchers need to make.[16] As in other nonparametric estimation, the cost of not knowing the parametric form might be the loss of efficiency in practice.

However, if we turn out to know the parametric form of the shock distribution, we can directly run Maximum Likelihood Estimation (MLE) for this SVAR-ZLB model. Specifying the non-Gaussian shock distribution, or at least flexibly approximating the non-Gaussian shock distribution, is a more practical option when we have a limited amount of data, as in most of the macroeconometrics literature. Although the likelihood function tend to behave differently when we specify different shock distributions, the point identification in the MLE is always guaranteed by this semiparametric identification scheme, which works for a generic non-Gaussian distribution.

With the above-mentioned two cases combined together, a more practical way of flexibly estimating the SVAR-ZLB model is to approximate the shock distribution with a mixture of normals, in the spirit of Sieve estimation. When we look at real data, we might not be able to determine which particular non-Gaussian shock distribution fits the data best. Nevertheless, the mixture-of-normal distribution is a flexible parametric distribution that can universally approximate any smooth shock distribution, when the number of mixture components goes to infinity (Goodfellow et al., 2016). In the limit, when the number of mixture components is infinity, we are back to the semiparametric estimation, in a way similar to using infinite basis functions in Sieve regression. In practice, using a data-driven method to determine the finite number of mixture components, we can closely approximate the unknown shock distribution with this mixture-of-normal parametric distribution and implement the model estimation in the parametric setup. A practical and efficient Bayesian inference routine is discussed in Section 1.6.

1.5 Generalize to Multivariate SVAR-ZLB Models

This section generalizes the bivariate model setup to the multivariate model setup and restate the identification arguments for the multivariate SVAR-ZLB model. In this section, I denote the n-dimensional observable vector as $y_t = (y_{1t}, y_{2t}')'$, where y_{1t} is the nominal interest rate and y_{2t} is a $(n-1)$-by-1 vector of

[16] The form of eigenvalue distribution will still hold if we average over two sets of points throughout (1.20) - (1.22). One can choose two sets of points that are likely to have significantly different eigenvalues to efficiently estimate A in practice.

private-sector variables, and I still use y_{1t}^* to represent the shadow interest rate.

The multivariate SVAR-ZLB model can similarly be written as a multivariate SVAR model with endogenously switched regimes, as in Section 1.2,

$$A(s_t)y_t(s_t) = Bx_t + \varepsilon_t \tag{1.34}$$

$$y_t^* = y_t(s_t) \tag{1.35}$$

$$y_{1t} = \max\{y_{1t}^*, 0\} \tag{1.36}$$

$$s_t = \mathbf{1}\{y_{1t} > 0\} \tag{1.37}$$

where s_t is the regime indicator, $y_t(s) = (y_{1t}(s), y_{2t}(s)')'$ is the latent outcome when fixed at the parameters in regime s, and $y_t^* = (y_{1t}^*, y_{2t}')'$ represents the shadow interest rate coupled with the observed private-sector variables. The lagged terms $x_t = (y_{t-1}', y_{t-2}', \cdots, y_{t-p}', 1)'$. The shocks are independent across time and across coordinates, and each follow the distribution f_i. The impact matrix across the two regimes have the following structure

$$A(1) = A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A(0) = A^* = \begin{bmatrix} A_{11} & A_{12} \\ A_{21}^* & A_{22} \end{bmatrix} \tag{1.38}$$

where the diagonal of A and A^* are normalized to be 1. We also impose the domain restriction on A and A^* as in (1.11) to guarantee the coherency. Note that in (1.38), A_{11} is a scalar, whereas A_{12}', A_{21}, A_{21}^* are $(n-1)$-by-1 vectors and A_{22} is a $(n-1)$-by-$(n-1)$ matrix.

Now we can restate the identification arguments that we show in Section 1.3 and Section 1.4.

Theorem 1.5. *Under Gaussian shocks, all the structural parameters, namely A, A_{21}^*, B in (1.34) - (1.38), will not be point-identified.*

See proof in Section 1.B.

Theorem 1.6. *Under non-Gaussian shocks, if all the regularity conditions in Theorem 1.2 - Theorem 1.4 hold, all the structural parameters, namely A, A_{21}^*, B in (1.34) - (1.38), will be point-identified.*

See proof in Section 1.B.

All the identification arguments in Section 1.3 and Section 1.4 are still valid, once we adjust the notations for the multivariate setup. First, Theorem 1.5 shows that we can still interpret the lack of point identification through the circular (or spherical) contour in the Gaussian likelihood. Second, Theorem 1.6 indicates that the generic semiparametric identification scheme still works for the multivariate case.

1.6 Bayesian Inference Routine and Simulation Study

This section proposes an efficient Bayesian inference routine for the SVAR-ZLB model and runs a simulation study to show whether the model is precisely estimated in practice. Section 1.6.1 discusses the advantage of Bayesian inference on this model when we use the data augmentation technique. Section 1.6.2 presents the setup and the result in the simulation study.

1.6.1 Bayesian Inference with a Gibbs Sampler

Using the data-augmentation and a Gibbs sampler in a Bayesian inference framework, we can easily deal with the censored nominal interest rate and handle the non-Gaussian shock distribution, which we approximate as a mixture of normals.

In practice, even if we do not know the parametric form of the shock distribution, we can still flexibly approximate that with a mixture-of-normal distribution, as mentioned in Section 1.4.4. For each shock ε_{it}, its distribution f_i is now approximated as

$$f_i \approx \sum_{k=1}^{\bar{K}} \pi_{ik} \mathcal{N}(\mu_{ik}, \sigma_{ik}^2) \tag{1.39}$$

where $k = 1, 2, \cdots, \bar{K}$ is the index of the mixture components. The mixture components are all independent across coordinates, to guarantee the independence across shocks. The number of components $\bar{K}$ needs to be specified in advance, but we can compare different specifications of $\bar{K}$ using Bayes factors after we run the Bayesian inference procedure. As a good starting point, this paper uses 4 mixture components to characterize the non-Gaussian shock distribution.[17]

When doing Bayesian inference, one big advantage we can rely on is the data augmentation. In the SVAR-ZLB model with a mixture-of-normal shock distribution, two latent variables will be extremely valuable for evaluating the likelihood if we can observe them in the data. The first one is Z_{it}, the component indicator in the mixture-of-normal distribution for the shock ε_{it}, which tell us which component is operative for generating that particular shock. Anther one is y_{1t}^*, the shadow interest rate, which is unconstrained interest rate before the censoring. In the Bayesian framework, the line between variables and parameters becomes vague. Thus, it is easy to treat Z_{it} and y_{1t}^* as the augmented data, and make posterior draws jointly on the augmented data and the model parameters in (1.7) and (1.8).

The lag order and the prior also need to be determined beforehand. First, the lag order can be determined by fitting a vanilla VAR model to the data and evaluating some well-known information criteria, such as Akaike information criterion (AIC). Second, the prior we choose is conjugate to the posterior for the structural parameters.[18] For example, the prior on B on μ_{ik} are independent Gaussian, the prior on $\{\pi_{ik}\}_{k=1}^{\bar{K}}$ is Dirichlet independent across i, and the prior on σ_{ik}^2 is independent inverse-Gamma. The details of the imposed prior for Bayesian inference can be found in 1.C.2.

The likelihood can be evaluated with or without the augmented data. On the one hand, when we do not involve the augmented data, (1.16) and (1.17) can yield the likelihood, where a one-dimensional numerical integral is recommended to compute (1.17) in practice.[19] This likelihood evaluation can help us find the posterior peak of all the parameters, without including the large dimension of the augmented data. On the other hand, when we have the augmented data, the likelihood with the observed data and the augmented

[17] The comparison of Bayes factors with different specifications of $\bar{K}$ will be added in the future version of this paper.

[18] The prior is conjugate to all the parameters except A and A^* in the Gibbs sampler.

[19] For the mixture-of-normal shock distribution, (1.17) can be computed analytically. However, the computation complexity will increase exponentially when we have more mixture components. Thus, using a one-dimensional numerical integral is more generic and robust in computation.

data given x_t will be

$$p(y_t, y_{1t}^*, \{Z_{it}\}_{i=1}^n | x_t) = \left(\prod_{i=1}^n p(Z_{it}|x_t) \right) p(y_t, y_{1t}^* | x_t, \{Z_{it}\}_{i=1}^n) \tag{1.40}$$

where

$$\begin{aligned} p(Z_{it} = k|x_t) &= \pi_{ik} \\ p(y_t, y_{1t}^*, s_t = 1|x_t, \{Z_{it}\}_{i=1}^n) &= \mathbf{1}\{y_{1t}^* = y_{1t}\} p(y_t, s_t = 1|x_t, \{Z_{it}\}_{i=1}^n) \\ p(y_t, y_{1t}^*, s_t = 0|x_t, \{Z_{it}\}_{i=1}^n) &= \mathbf{1}\{y_{1t}^* < 0\} p((y_{1t}^*, y_{2t}), s_t = 0|x_t, \{Z_{it}\}_{i=1}^n) \end{aligned}$$

Note that the last two formulas are simply evaluating the likelihood of an unconstrained linear SVAR model with Gaussian shocks.

After we combine the prior with the likelihood, a Gibbs sampler can be applied. The Gibbs sampler utilizes the tractable posterior of a subset of parameters conditioning on other parameters and samples from the posterior alternatingly for the different sets of parameters. In particular, each time we condition on other parameters and make posterior draws on a subset of the parameters, in the order: (i) Z_{it}; (ii) π_{ik}; (iii) B and μ_{ik}; (iv) σ_{ik}^2; (v) A and A^*; (vi) y_{1t}^*. To compute the posterior density, we use the Gibbs sampler to generate 10,000 draws from the posterior. The details of the Gibbs sampler can be found in Section 1.C.

1.6.2 Simulation Study

I specify a trivariate SVAR-ZLB model, calibrated towards the DGP of real data, with a mixture-of-normal shock distribution. I then run Bayesian inference based on Section 1.6.1 to see if a point-identified model will be precisely estimated in practice.

The simulation DGP is calibrated towards the real data in Section 1.7. In particular, I calibrate A, A^*, B to match the persistence of the real data and the simultaneous equation structure. I also use the calibrated mixture-of-normal distribution to mimic the non-Gaussian distribution in the data, where the number of mixture components is 4. In addition, the lag order of the simulation DGP is also calibrated to the lag structure of the real data, which is set to 4 and treated as known in the simulation. I then start the simulation at the sample mean of the real data and then simulate for 421 periods, which has exactly the same length of my real data. Since the nominal interest rate turns out to be relatively persistent, the simulated ZLB occurrence can be too high or too low compared to the real data. Thus, I tune the random seed in the simulation to pick one realization of the simulated path that match the fraction of ZLB periods in the real data, which is about 20% of the sample period. The quantiles of the posterior density for A and A^* are reported in Table 1.1.[20]

The credible intervals demonstrate that this model, which is pointed identified through non-Gaussian shocks, can be precisely estimated in practice. As shown in Table 1.1, the posterior density of each parameter is around the true parameter value. The 68% credible intervals (from 16%-quantile to 84%-quantile) and the 90% credible intervals (from 5%-quantile to 95%-quantile) all have the reasonable widths and still well

[20]The credible intervals in the simulation study for all the other parameters will be listed in an online supplement appendix in the future version.

Quantiles of Posterior Density in Simulation

	5%	16%	50%	84%	95%	True
A_{21}	-0.17	-0.11	-0.00	0.12	0.18	-0.12
A_{31}	-0.17	-0.12	-0.04	0.04	0.09	0.01
A_{12}	-0.24	-0.18	-0.02	0.13	0.20	0.03
A_{32}	-0.02	0.01	0.10	0.19	0.29	0.02
A_{13}	-0.28	-0.24	-0.09	0.05	0.13	-0.17
A_{23}	-0.42	-0.29	-0.16	-0.05	0.01	-0.02
A^*_{21}	-0.52	-0.36	-0.15	-0.01	0.06	-0.27
A^*_{31}	-0.18	-0.11	0.03	0.18	0.30	-0.02

Table 1.1: The first five columns represent the quantiles of the posterior density of each parameter. The last column represents the true parameter value in the simulation. The diagonal of A and A^* is normalized to be 1.

capture the true values. With this realistic simulation setup, we show that this model can be precisely estimated, given the sample size is reasonably large and the occurrence of ZLB periods is not extremely rare.

1.7 Empirical Results

This section uses the proposed efficient Bayesian inference routine to estimate the SVAR-ZLB model using real data. By fitting the model to three key US economic variables, we find that the effect of unconventional monetary policy is small and short-lived. Moreover, the fat-tail of the shocks in the estimated model support the non-Gaussianity of shocks in reality. The implied shadow interest rate stays small in magnitude in the ZLB periods, except a sudden decline at the beginning of the Great Recession.

This paper uses three monthly US time series to run the empirical analysis. The monthly dataset includes the federal fund rate, the output growth, and the inflation rate.[21] The sample period I choose is 1985:Jan - 2020:Jan, which is usually considered as a moderate period without too much variation in the shock variance.[22] The time series of these three US macroeconomic variables are plotted in Figure 1.3. The ZLB periods span from 2008:Dec to 2015:Dec.[23] The Bayesian inference setup is the same as in Section 1.6.2, where the lag order is set to be 4 based on AIC, the number of mixture components is set to be 4 (symmetric on both sides of zero), and the conjugate prior is imposed.

When we estimate the model using the efficient Bayesian inference routine, we find strong evidence for the non-Gaussianity of the shocks. Using the posterior modal estimates of the parameters, we can characterize the shock distribution with a QQ plot as in Figure 1.4. Each shock exhibits a fat-tail in the distribution, which would help the model identification as mentioned in Section 1.3.3.[24] Moreover, the estimated mixture-

[21] The output growth is defined as log difference of industrial production, and the inflation rate is defined as the log difference of PCE price level. The data source is Fred. See the data labels of "FEDFUNDS", "INDPRO" and "PCEPI" for details.

[22] Starting from the moderate period in the empirical study is usually a challenge for identifying the monetary policy shocks, since the results are found to be different from the ones using the data back to 1970s (Ramey, 2016). However, this paper tries to rely on the non-Gaussianity of shocks to identify the monetary policy shocks in the moderate period.

[23] In the ZLB periods, the effective federal funds rate is clearly constrained at zero, but not exactly zero. To fit the sharp censoring at zero in the model, this paper around all the small values from 0 to 25 basis points to zero. Researchers can try to model this small movements from 0 to 25 basis points in future work.

[24] Specifying a mixture-of-normal distribution does not necessarily imply a fat-tail. The fat tail is an estimate through the data.

Time Series of Macroeconomic Variables

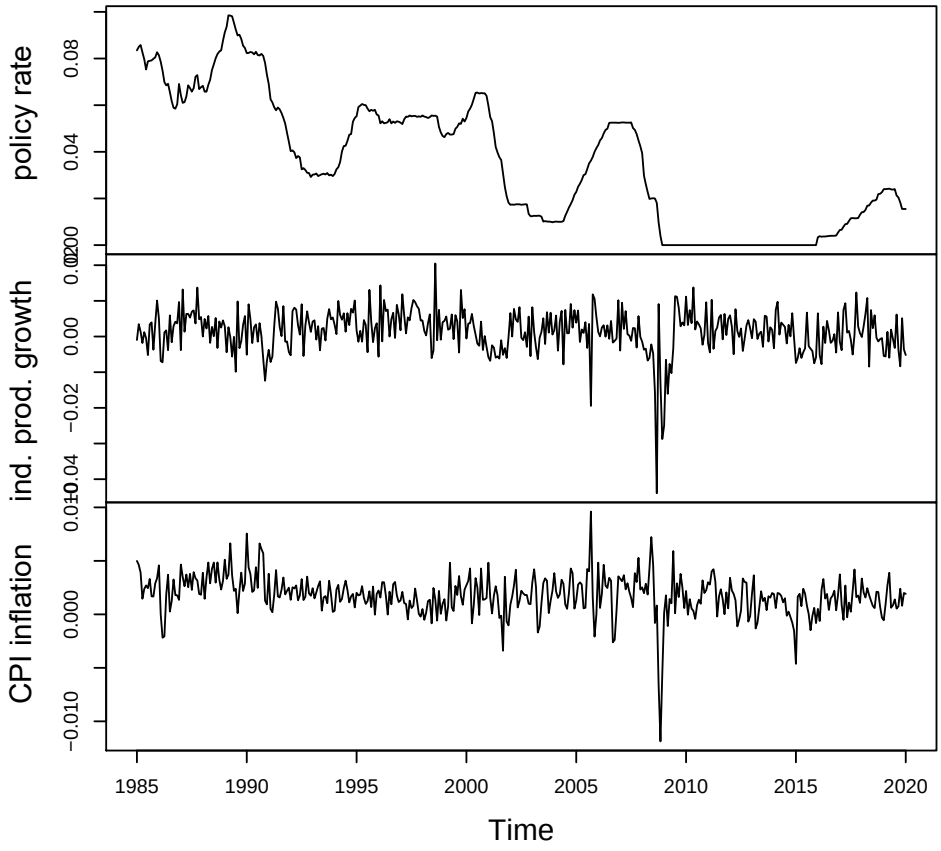
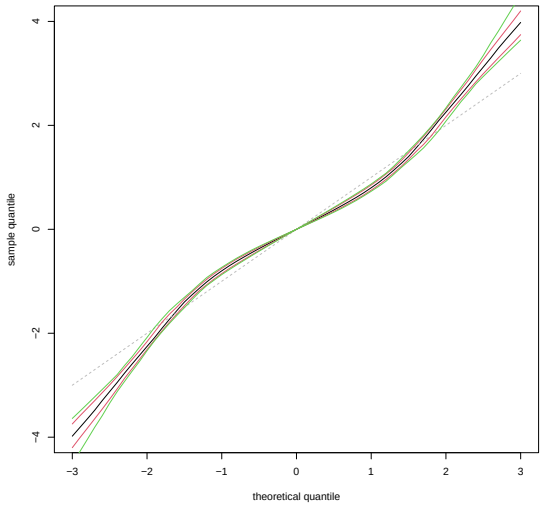


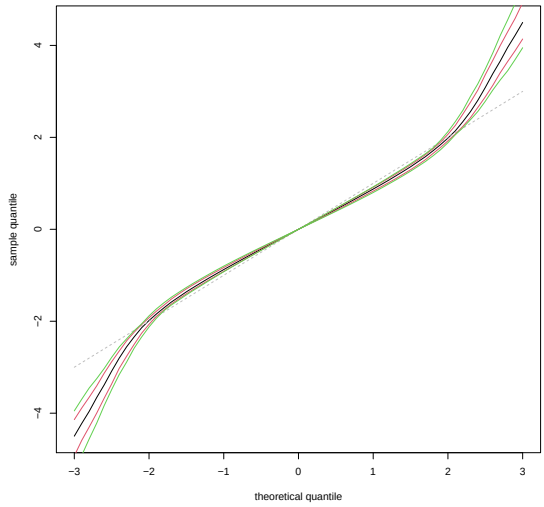
Figure 1.3: The sample period is 1985:Jan to 2020:Jan. For the variables, "policy rate" means the federal funds rate, "ind. prod. growth" means the output growth measured via industrial production, "CPI inflation" means the inflation rate measured via the PCE price level. The unit is percentage points.

QQ Plot for Shock Distribution

monetary shock



output shock



price shock

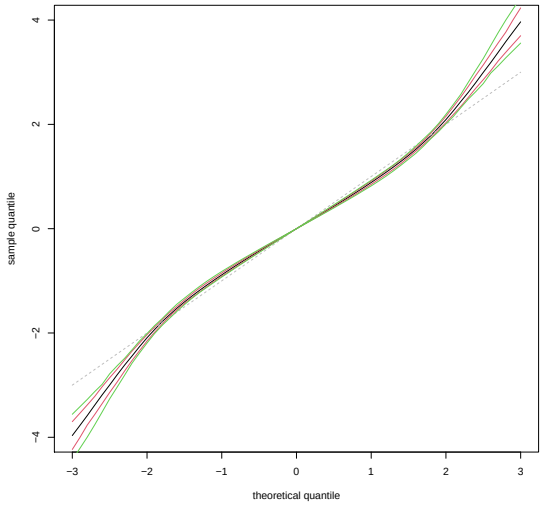


Figure 1.4: The horizontal axis is the quantiles of Gaussian distribution, and the vertical axis is the quantiles of the shock distribution.

Quantiles of Posterior Density in Empirical Analysis

	5%	16%	50%	84%	95%
A_{21}	-0.14	-0.11	-0.06	-0.03	-0.00
A_{31}	-0.05	-0.03	0.02	0.07	0.09
A_{12}	-0.12	-0.08	-0.03	0.02	0.06
A_{32}	-0.12	-0.05	0.05	0.15	0.24
A_{13}	-0.29	-0.25	-0.16	-0.08	-0.03
A_{23}	-0.31	-0.21	-0.07	0.04	0.17
A^*_{21}	-0.33	-0.26	-0.15	-0.00	0.10
A^*_{31}	-0.13	-0.05	0.07	0.19	0.28

Table 1.2: The first five columns represent the quantiles of the posterior density of each parameter. The diagonal of A and A^* is normalized to be 1.

of-normal shock distribution tends to have a high probability on one component with a small variance, while having a low probability on other components with a big variance. The normal mean of these mixture components all tend to be close to zero in the posterior.

The posterior quantiles for all the parameters are reported in Table 1.2.[25] We notice that the off-diagonal elements in A and A^* tend to be intermediate, which coincides with the intermediate correlation across the three variables. The short-run effects of both conventional and unconventional monetary policy tend to be small, and not significantly from zero. The monetary equation shows a persistent lag structure in B, which explains the persistence in the federal funds rate. All the shocks have at least one mixture component with a big variance, which contributes to the fat tail in the shock distribution and helps the model identification.

Given the posterior density of all the parameters, we can also plot the impulse responses of the three economic variables, in the standard regime and the ZLB regime respectively. In Figure 1.5, we use the parameters in the standard regime and trace out the impulse responses, where as in Figure 1.6 we use the parameters in the ZLB regime to achieve the same goal. The three shocks, by examining their effects on the economic variables, can be labeled as the monetary policy shock, the demand shock, and the supply shock. Figure 1.5 shows the impulse response to a one-standard-deviation shock when we stay in the standard regime. The monetary policy shock increases the nominal interest rate and has a persistent effect, and the negative response in output growth and inflation rate is also similar to conventional wisdom.[26] In contrary, Figure 1.6 presents the similarly defined responses in the ZLB regime. Now the monetary policy shock can not persistently affect the shadow interest rate since the increase in the shadow interest rate cannot pass to the nominal interest rate in the next period. Thus, the long-run effect of the unconventional monetary policy is quite limited.[27] The empirical results suggest the unconventional monetary policy has a small and short-lived effect, compared to the conventional monetary policy.

We can now easily discuss the posterior distribution of the shadow interest rate, as shown in Figure 1.7. By using the data augmentation technique in the efficient Bayesian inference routine, it is easy to draw from

[25] The credible intervals in the empirical analysis for all the other parameters will be listed in an online supplement appendix in the future version.

[26] We also see a little evidence of the price puzzle, which is a common issue in the monetary SVAR literature (Ramey, 2016).

[27] A possible way to allow a persistent effect in the shadow interest rate is to add lagged shadow interest rates in the model. Further research is useful to consider the persistence of the shadow interest rate.

Impulse Response in Standard Regime

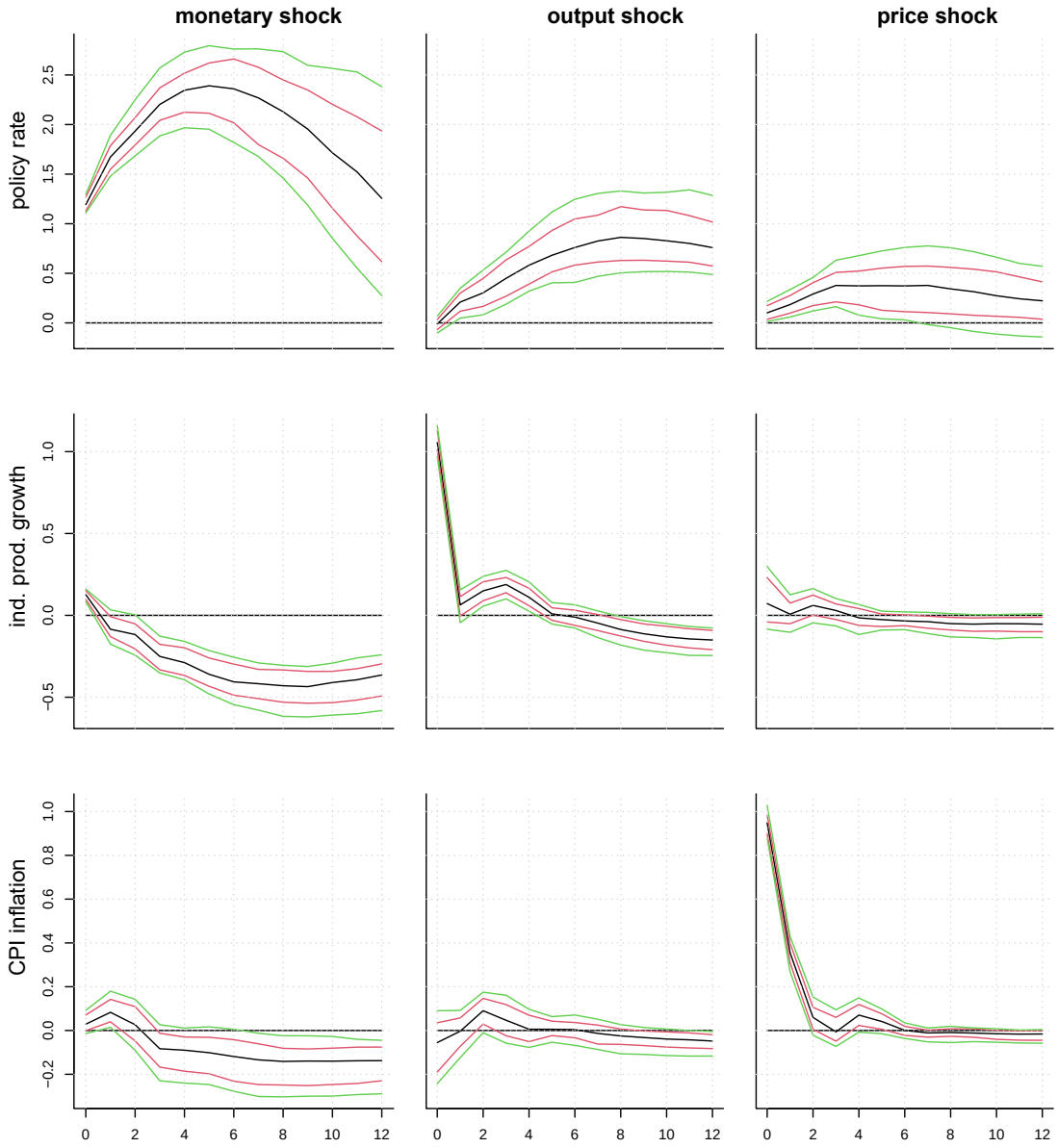


Figure 1.5: Impulse response w.r.t. one-standard deviation shock. Each row represent one variable, and each column represent one shock. Note the first row represents the response of the nominal interest rate. The 68% and 90 % credible bands are plotted in red and green. The horizontal axis spans from 0 to 12 months after the shock occurrence.

Impulse Response in ZLB Regime

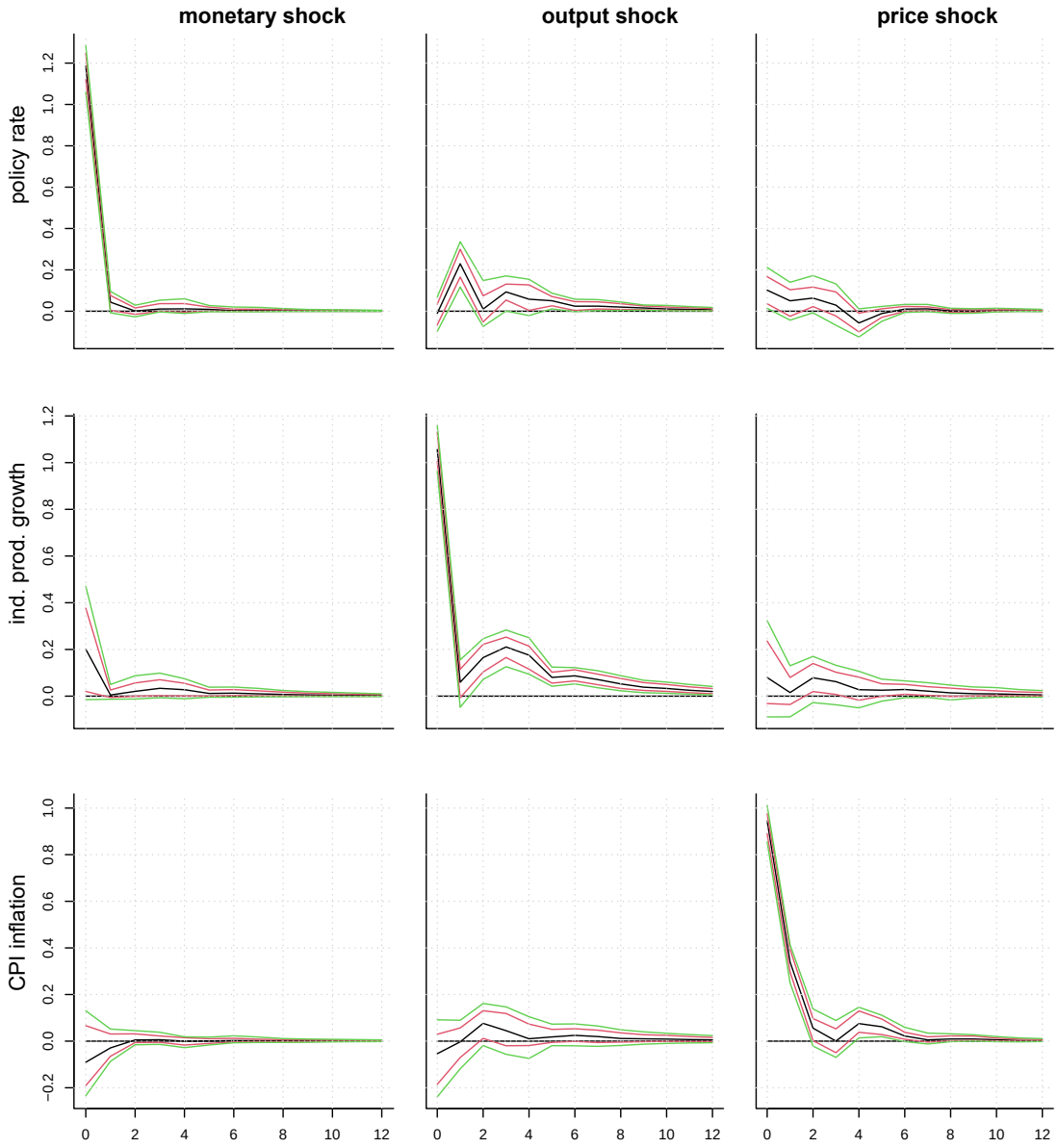


Figure 1.6: Impulse response w.r.t. one-standard deviation shock. Each row represent one variable, and each column represent one shock. Note the first row represents the response of the shadow interest rate. The black line is the posterior mean, the red line is the 68% credible interval and the green line is the 90% credible interval. The horizontal axis spans from 0 to 12 months after the shock occurrence.

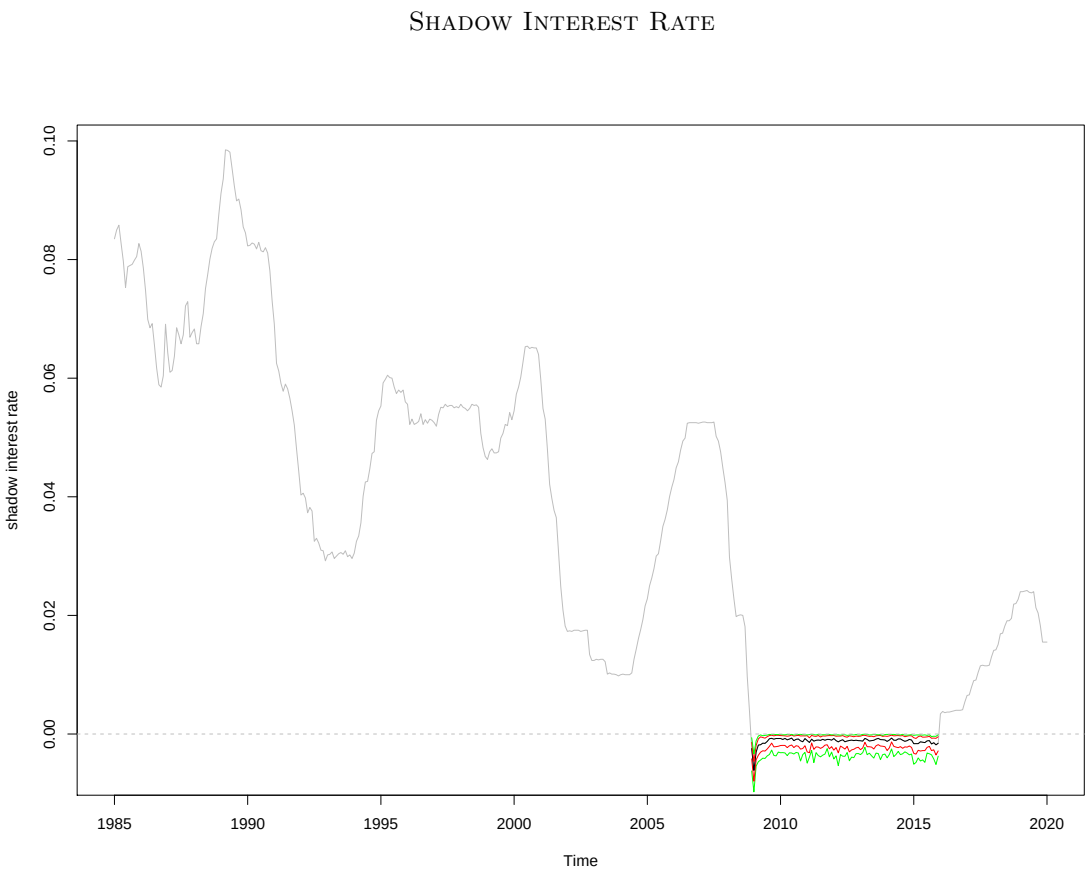


Figure 1.7: Negative shadow interest rate between 2008:Dec and 2015:Dec. The gray line is the nominal interest rate. The black line, the red line, and the green line are the posterior mean, the 68% credible interval and the 90% credible interval.

the posterior of y_{1t}^* in the ZLB periods. The overall magnitude of the shadow interest rate is small, meaning the desired interest rate in central banks was not largely deviating from zero in the ZLB periods, except the sudden decline at the beginning of the Great Recession.

1.8 Conclusion

This paper uses a SVAR-ZLB model to characterize the censored nominal interest rate and the effect of unconventional monetary policy, and considers the model identification when we do not assume the short-run effect of unconventional monetary policy is zero. I first interpret why we cannot achieve point identification of this SVAR-ZLB model if the shocks are Gaussian, from the new likelihood perspective. The Gaussian shocks have a circular contour in the likelihood function, so any rotated version of the model will fit the censored and the uncensored data equally well. It turns out this lack of point identification is only a problem for the Gaussian distribution. If we remove the Gaussian assumption and specify the empirically relevant non-Gaussian distribution for the shocks, the non-Gaussian shocks will help point-identify the model. Then I propose a generic semiparametric identification scheme, following independent component analysis and Heckman selection models, to prove the point identification. Using ICA to identify the structural parameters will encounter a non-linearity problem from the censoring, and I adjust the ICA techniques through Hessian matrices of log densities to deal with this problem. Moreover, I design an efficient Bayesian inference routine to facilitate the model estimation in practice. The empirical results show that the unconventional monetary policy has a small and transitory effect.

There are also interesting questions that need to be answered in the future research. First, when we specify the non-Gaussian shock distribution, we might ask if we still have consistency in MLE when we slightly misspecify the non-Gaussian distribution. For example, we might know the shock is t-distributed but we might get a wrong estimate for the degree of freedom. Second, although we have the model being point-identified, we might wonder if the model has weak identification issues in finite samples. It is still unknown whether the model identification from ICA and semiparametric Heckman selection models is efficient in finite samples.

Appendix 1.A Coherency Conditions on DGP

We assume the coherency of DGP, which implies the continuity of variables w.r.t. shocks and the existence of a unique equilibrium, for any shock value. The two coherency conditions we impose on the DGP are continuity and uniqueness. As in Aruoba et al. (2021), we need continuity as a necessary condition to achieve coherency, and the continuity and the uniqueness together are necessary and sufficient conditions for coherency. We now explain how we can guarantee these two conditions in the SVAR-ZLB model.

For continuity, it is guaranteed by the model setup where the two regimes only differ in A_{21} and A_{21}^*. First, A_{11} cannot change across regimes, because that will make the definition of the shadow interest rate vague and the effect of unconventional monetary policy will be not properly defined. For example, if we make A_{11} half as much as before in the ZLB regime, then the accordingly defined shadow interest rate will be twice as large as before. Thus the effect of unconventional monetary policy will also be half as much as before. Second, A_{12}, A_{22} and B cannot be contingent on the regime, because the contingency in these parameters will make the y_{2t} not continuous across the two regimes when $y_{1t}^* = 0$. Notice that, in the marginal case of $y_{1t}^* = 0$, the value of y_{2t} follows

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} 0 \\ y_{2t} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21}^* & A_{22} \end{bmatrix} \begin{bmatrix} 0 \\ y_{2t} \end{bmatrix} = \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} y_{2t} = Bx_t + \varepsilon_t \tag{1.A.1}$$

Hence, the same value of y_{2t} will realize no matter we are using the standard regime or the ZLB regime, when $y_{1t}^* = 0$. With the only change arising in A_{21} and A_{21}^*, we can conclude that, given x_t, y_{1t}^* and y_{2t} will continuously change with ε_t.

As for uniqueness, we need the domain restriction, i.e. $|A|/|A^*| > 0$, to guarantee the existence of a unique equilibrium. We can apply (Gouriéroux et al. (1980), Theorem 1), to get the necessary and sufficient condition in my model setup, which is $\text{sign}(|A|) = \text{sign}(|A^*|)$.

Appendix 1.B Proofs

1.B.1 Proofs of Lemmas

Lemma 1.B.1. *For any continuous shock distribution f_i, the likelihood evaluation formulas in the y-space and in the ε-space are equivalent. In particular, for the standard regime, we can show that (1.16) is equivalent to imputing the shocks $\varepsilon_t = Ay_t - Bx_t$ and evaluating*

$$p(y_t, s_t = 1|x_t) = |A| \cdot \mathbf{f}(\varepsilon_t) \tag{1.B.1}$$

whereas for the ZLB regime, we can show that (1.17) is equivalent to imputing the pseudo shocks $\varepsilon_t^0 = A^*(0, y_{2t})' - Bx_t$, and integrating out the true shocks subject to a linear constraint

$$p(y_t, s_t = 0|x_t) = \int_{-\infty}^{\varepsilon_{1t}^0} f(\varepsilon_{1t}) f\left(\varepsilon_{2t}^0 - \frac{G_{21}^*}{G_{22}^*}(\varepsilon_{1t} - \varepsilon_{1t}^0)\right) \frac{1}{G_{22}^*} d\varepsilon_{1t} \tag{1.B.2}$$

Similarly, the full likelihood evaluated in the ε-space is comprised of (1.B.1) and (1.B.2), as written in (1.18).

Proof. The equivalence between (1.16) is equivalent to (1.B.1) can be easily seen once we plug $\varepsilon_t = Ay_t - Bx_t$.

To prove the equivalence between (1.17) is equivalent to (1.B.2), we can start from the y-space:

$$\begin{aligned}
&\int_{-\infty}^{0} |A^*| f_1(A_{1\cdot}(y_{1t}^*, y_{2t})' - B_{1\cdot}x_t) f_2(A_{2\cdot}^*(y_{1t}^*, y_{2t})' - B_{2\cdot}x_t) dy_{1t}^* \\
=&\int_{-\infty}^{0} f(A_{1\cdot}(y_{1t}^*, y_{2t})' - B_{1\cdot}x_t) \cdot \\
&f\left(A_{22}y_{2t} - B_{2\cdot}x_t + \frac{A_{21}^*}{A_{11}}[(A_{11}y_{1t}^* + A_{12}y_{2t} - B_{1\cdot}x_t) - (A_{12}y_{2t} - B_{1\cdot}x_t)]\right) \frac{|A^*|}{A_{11}} A_{11} dy_{1t}^* \\
=&\int_{-\infty}^{\varepsilon_{1t}^0} f(\varepsilon_{1t}) f\left(\varepsilon_{2t}^0 + \frac{A_{21}^*}{A_{11}}(\varepsilon_{1t} - \varepsilon_{1t}^0)\right) \frac{|A^*|}{A_{11}} d\varepsilon_{1t} \\
=&\int_{-\infty}^{\varepsilon_{1t}^0} f(\varepsilon_{1t}) f\left(\varepsilon_{2t}^0 - \frac{G_{21}^*}{G_{22}^*}(\varepsilon_{1t} - \varepsilon_{1t}^0)\right) \frac{1}{G_{22}^*} d\varepsilon_{1t} \qquad (1.B.3)
\end{aligned}$$

□

Remark. From Lemma 1.B.1, we can explicitly characterize the different lines in the right panel of Figure 1.1. The boundary line (l_b) represents $0 = \beta_{1\cdot}x_t + G_{1\cdot}\varepsilon_t$. The integral line ($l_g$) represents $y_{2t} = \beta_{2\cdot}^* + G_{2\cdot}^*\varepsilon_t$. Note that the pseudo shocks by definition are also on the integral line (l_g), i.e. $y_{2t} = \beta_{2\cdot}^* x_t + G_{2\cdot}^*\varepsilon_t^0$, and thus the integral line (l_g) can be equivalently written as $\varepsilon_{2t} = \varepsilon_{2t}^0 - \frac{G_{21}^*}{G_{22}^*}(\varepsilon_{1t} - \varepsilon_{1t}^0)$, which shows up in the last step of (1.B.3).

Lemma 1.B.2. *The model can be normalized as with $\mathbb{A}, \mathbb{A}^*, \mathbb{B}$, with the shock η_{it} whose variance is 1.*

Proof. The standard regime can be written by dividing by σ_1 and σ_2 for the two rows separately,

$$\begin{bmatrix} A_{11}/\sigma_1 & A_{12}/\sigma_1 \\ A_{21}/\sigma_2 & A_{22}/\sigma_2 \end{bmatrix} \begin{bmatrix} y_{1t}^* \\ y_{2t} \end{bmatrix} = \begin{bmatrix} B_{1\cdot}/\sigma_1 \\ B_{2\cdot}/\sigma_2 \end{bmatrix} x_t + \begin{bmatrix} \varepsilon_{1t}/\sigma_1 \\ \varepsilon_{2t}/\sigma_2 \end{bmatrix} \qquad (1.B.4)$$

whereas in the ZLB regime,

$$\begin{bmatrix} A_{11}/\sigma_1 & A_{12}/\sigma_1 \\ A_{21}^*/\sigma_2 & A_{22}/\sigma_2 \end{bmatrix} \begin{bmatrix} y_{1t}^* \\ y_{2t} \end{bmatrix} = \begin{bmatrix} B_{1\cdot}/\sigma_1 \\ B_{2\cdot}/\sigma_2 \end{bmatrix} x_t + \begin{bmatrix} \varepsilon_{1t}/\sigma_1 \\ \varepsilon_{2t}/\sigma_2 \end{bmatrix} \qquad (1.B.5)$$

Hence, we can define the re-normalized parameters to simplify the notation

$$\mathbb{A} = \begin{bmatrix} A_{11}/\sigma_1 & A_{12}/\sigma_1 \\ A_{21}/\sigma_2 & A_{22}/\sigma_2 \end{bmatrix} \qquad (1.B.6)$$

$$\mathbb{A}^* = \begin{bmatrix} A_{11}/\sigma_1 & A_{12}/\sigma_1 \\ A_{21}^*/\sigma_2 & A_{22}/\sigma_2 \end{bmatrix} \qquad (1.B.7)$$

$$\mathbb{B} = \begin{bmatrix} B_{1\cdot}/\sigma_1 \\ B_{2\cdot}/\sigma_2 \end{bmatrix} \qquad (1.B.8)$$

$$\eta_t = \begin{bmatrix} \varepsilon_{1t}/\sigma_1 \\ \varepsilon_{2t}/\sigma_2 \end{bmatrix} \qquad (1.B.9)$$

We can see (1.B.4) and (1.B.5) is an equivalent way to specify the SVAR-ZLB model, except the shock variance is now 1.

The structure of this SVAR-ZLB model still requires that $\mathbb{A}$ and $\mathbb{A}^*$ have the same element except the bottom-left corner. □

Lemma 1.B.3. *Given the true value of $\mathbb{A}$, $\mathbb{A}^*$, and a rotation matrix R, there exists a unique $\tilde{\mathbb{A}}^*$ such that the model setup is still valid and $\tilde{\mathbb{G}}^*_{2\cdot} = \mathbb{G}^*_{2\cdot}R'$. The mapping rule $R^* : (\mathbb{A}, \mathbb{A}^*, R) \to \tilde{\mathbb{A}}^*$ is stated as*

$$\tilde{\mathbb{A}}^* = \begin{bmatrix} (R\mathbb{A})_{11} & (R\mathbb{A})_{12} \\ \frac{(R\mathbb{A})_{11}}{(R\mathbb{A}^*)_{11}}(R\mathbb{A}^*)_{21} & (R\mathbb{A})_{22} \end{bmatrix} \tag{1.B.10}$$

Proof. With A and R, we can first define the rotated matrix $\tilde{A} = RA$. For the new value of $\tilde{A}^*$, we first need to force $\tilde{A}^*$ and $\tilde{A}$ have the same elements except for the bottom-left corner, to preserve the structure of SVAR-ZLB model. Thus $\tilde{A^*}_{i,j} = (R \cdot A)_{i,j}$, for $(i,j) = (1,1),(1,2),(2,2)$. Now the model setup is preserved. We further need to find a new value $\tilde{A}^*_{21}$ in $\tilde{A}^*$, such that $\tilde{G}^*_{2\cdot} = G^*_{2\cdot}R'$, where $\tilde{G}^* = \tilde{A}^{*-1}$.

We can first show (1.B.10) is a sufficient condition for $\tilde{G}^*_{2\cdot} = G^*_{2\cdot}R'$. By using the $\tilde{A}^*$ generated by the mapping (1.B.10), we have

$$\begin{aligned}
\tilde{G}^*_{2\cdot} &= \begin{bmatrix} \tilde{G}^*_{21} & \tilde{G}^*_{22} \end{bmatrix} \\
&= \frac{1}{\tilde{A}^*_{11}\tilde{A}^*_{22} - \tilde{A}^*_{12}\tilde{A}^*_{21}} \begin{bmatrix} -\tilde{A}^*_{21} & \tilde{A}^*_{11} \end{bmatrix} \\
&= \frac{1}{(RA)_{11}(RA)_{22} - (RA)_{12}\frac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{21}} \begin{bmatrix} -\frac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{21} & (RA)_{11} \end{bmatrix} \\
&= \frac{1}{\frac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{11}(RA^*)_{22} - (RA^*)_{12}\frac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{21}} \begin{bmatrix} -\frac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{21} & \frac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{11} \end{bmatrix} \\
&= \frac{1}{(RA^*)_{11}(RA^*)_{22} - (RA^*)_{12}(RA^*)_{21}} \begin{bmatrix} -(RA^*)_{21} & (RA^*)_{11} \end{bmatrix} \\
&= \frac{1}{|RA^*|} \begin{bmatrix} -(RA^*)_{21} & (RA^*)_{11} \end{bmatrix} \\
&= e_2'(RA^*)^{-1} \\
&= e_2'(G^*R') \\
&= G^*_{2\cdot}R'
\end{aligned} \tag{1.B.11}$$

Then, let's prove that (1.B.10) is a necessary condition for $\tilde{G}^*_{2\cdot} = G^*_{2\cdot}R'$. Given $\tilde{A^*}_{i,j} = (R \cdot A)_{i,j}$, for $(i,j) = (1,1),(1,2),(2,2)$, we need $\tilde{A}^*_{21}$ to satisfy

$$e_2' \begin{bmatrix} (RA)_{11} & (RA)_{12} \\ \tilde{A}^*_{21} & (RA)_{22} \end{bmatrix}^{-1} = G^*_{2\cdot}R' \tag{1.B.12}$$

where $\tilde{A}^*_{21}$ is the only unknown and the number of equations is two. Hence, there are at most one $\tilde{A}^*_{21}$ to make (1.B.12) hold. □

REMARK: It is noteworthy that in this SVAR model with ZLB, we also consider the case when we rotate the true A and B matrix by θ to get a new set of parameter values, namely $\tilde{A} = R \cdot A$ and $\tilde{B} = R \cdot B$. Since the level curve of joint standard normal is a circle. If we can rotate everything in the shock space by θ, we will preserve the likelihood. First the boundary line will rotate by θ, because the boundary line is now $\tilde{G}_{1\cdot}\tilde{m} + \tilde{G}_{1\cdot}\tilde{\varepsilon}_t = 0$. Second, the uncensored observation will rotate by θ, i.e. $\tilde{\varepsilon}_t = R \cdot \varepsilon_t$. Third, the censored observation will also rotate, because they can be treated as the marginal cases in the standard regime. Finally, I will prove that we can rotate the integral line by θ.

1.B.2 Proof of Theorems

Proof of Theorem 1.1.

Proof. Under the true parameter values, we are going to use the following representation

$$\mathbb{A}y_t(1) = \mathbb{B}x_t + \eta_t \tag{1.B.13}$$

$$\mathbb{A}^* y_t(0) = \mathbb{B}x_t + \eta_t \tag{1.B.14}$$

Note that $\eta_t \overset{\text{iid}}{\sim} \mathcal{N}(0, I)$, and thus given x_t we have

$$\begin{bmatrix} y_{1t}(1) \\ y_{2t}(1) \end{bmatrix} \sim \mathcal{N}(\beta x_t, \mathbb{G}\mathbb{G}') \tag{1.B.15}$$

$$\begin{bmatrix} y_{1t}(1) \\ y_{2t}(0) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_{1\cdot} \\ \beta^*_{2\cdot} \end{bmatrix} x_t, \begin{bmatrix} \mathbb{G}_{1\cdot} \\ \mathbb{G}^*_{2\cdot} \end{bmatrix} \begin{bmatrix} \mathbb{G}_{1\cdot} \\ \mathbb{G}^*_{2\cdot} \end{bmatrix}' \right) \tag{1.B.16}$$

To evaluate the likelihood in the standard regime,

$$p(y_t, s_t = 1|x_t) = p(y_t(1)|x_t) = \phi(y_t(1); \beta x_t, \mathbb{G}\mathbb{G}') \tag{1.B.17}$$

whereas the likelihood in the ZLB regime,

$$\begin{aligned} p(y_t, s_t = 0|x_t) =& p(y_{1t}(1) < 0, y_{2t}(0)|x_t) \\ =& p(y_{2t}(0)|x_t) \cdot p(y_{1t}(1) < 0|y_{2t}(0), x_t) \\ =& \phi(y_{2t}(0); \beta^*_{2\cdot}x_t, \mathbb{G}^*_{2\cdot}\mathbb{G}^{*\prime}_{2\cdot}) \cdot \\ & \Phi(0; \beta_{1\cdot}x_t + \Omega_{12}\Omega_{22}^{-1}(y_{2t}(0) - \beta^*_{2\cdot}x_t), \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}) \end{aligned} \tag{1.B.18}$$

where $\phi(\cdot; m, \tau^2)$ means Gaussian density with mean m and variance τ^2, and $\Phi(\cdot; m, \tau^2)$ means the corresponding Gaussian distribution function, and

$$\Omega = \begin{bmatrix} \mathbb{G}_{1\cdot} \\ \mathbb{G}^*_{2\cdot} \end{bmatrix} \begin{bmatrix} \mathbb{G}_{1\cdot} \\ \mathbb{G}^*_{2\cdot} \end{bmatrix}'$$

Now if we rotate the true parameter values to get $\tilde{\mathbb{A}} = R\mathbb{A}$, $\tilde{\mathbb{B}} = R\mathbb{B}$, and $\tilde{\mathbb{A}}^* = R^*(\mathbb{A}^*; \mathbb{A}, R)$. Then the

model representation is

$$\tilde{\mathbb{A}} y_t(1) = \tilde{\mathbb{B}} x_t + \tilde{\eta}_t = \tilde{\mathbb{A}}^* y_t(0) \tag{1.B.19}$$

Note that now $\tilde{\mathbb{G}} = \mathbb{G}R'$, $\tilde{\mathbb{G}}^*_{2\cdot} = \mathbb{G}^*_{2t}R'$, and thus

$$\tilde{\mathbb{G}}\tilde{\mathbb{G}}' = \mathbb{G}\mathbb{G}' \tag{1.B.20}$$

$$\begin{bmatrix} \tilde{\mathbb{G}}_{1\cdot} \\ \tilde{\mathbb{G}}^*_{2\cdot} \end{bmatrix} \begin{bmatrix} \tilde{\mathbb{G}}_{1\cdot} \\ \tilde{\mathbb{G}}^*_{2\cdot} \end{bmatrix}' = \begin{bmatrix} \mathbb{G}_{1\cdot} \\ \mathbb{G}^*_{2\cdot} \end{bmatrix} \begin{bmatrix} \mathbb{G}_{1\cdot} \\ \mathbb{G}^*_{2\cdot} \end{bmatrix}' \tag{1.B.21}$$

$$\tilde{\mathbb{G}}\tilde{\mathbb{B}} = \mathbb{G}\mathbb{B} = \beta \tag{1.B.22}$$

$$\tilde{\mathbb{G}}^*_{2\cdot}\tilde{\mathbb{B}} = \mathbb{G}^*_{2\cdot}\mathbb{B} = \beta^*_{2\cdot} \tag{1.B.23}$$

So the distribution of $y_t(1)$ and $(y_{1t}(1), y_{2t}(0)$ is the same as (1.B.15) (1.B.16). Hence the likelihood value is the same after we use the rotated version of parameters. □

Proof of Theorem 1.5.

Proof. The proof is mostly the same as the proof in Theorem 1.1, because we can simply change the notations from the bivariate case to the multivariate case to complete most of the proof in Theorem 1.1. However, the only part that we cannot be trivially complete when we change the notations is the proof in Lemma 1.B.3.

We can still use the mapping rule (1.B.10) and show it is a sufficient condition for $\tilde{G}^*_{2\cdot} = G^*_{2\cdot}R'$ in this multivariate case. By using the matrix inverse formula a 2-by-2 block matrix and the mapping rule, we get

$$\begin{aligned}
\tilde{G}^*_{2\cdot} &= \left[-\left(\tilde{A}^*_{22} - \tilde{A}^*_{21}\tilde{A}^{*-1}_{11}\tilde{A}^*_{12}\right)^{-1} \tilde{A}^*_{21}\tilde{A}^{*-1}_{11}, \quad \left(\tilde{A}^*_{22} - \tilde{A}^*_{21}\tilde{A}^{*-1}_{11}\tilde{A}^*_{12}\right)^{-1} \right] \\
&= \left[-\left((RA)_{22} - \tfrac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{21}(RA)^{-1}_{11}(RA)_{12}\right)^{-1} \tfrac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{21}(RA)^{-1}_{11}, \quad \left((RA)_{22} - \tfrac{(RA)_{11}}{(RA^*)_{11}}(RA^*)_{21}(RA)^{-1}_{11}(RA)_{12}\right)^{-1} \right] \\
&= \left[-\left((RA^*)_{22} - (RA^*)_{21}(RA^*)^{-1}_{11}(RA^*)_{12}\right)^{-1} (RA^*)_{21}(RA^*)^{-1}_{11}, \quad \left((RA^*)_{22} - (RA^*)_{21}(RA^*)^{-1}_{11}(RA^*)_{12}\right)^{-1} \right] \\
&= e_2'(RA^*)^{-1} \\
&= e_2'(G^*R') \\
&= G^*_{2\cdot}R'
\end{aligned} \tag{1.B.24}$$

Then, we can still prove that (1.B.10) is a necessary condition for $\tilde{G}^*_{2\cdot} = G^*_{2\cdot}R'$. Given $\tilde{A}^*{}_{i,j} = (R \cdot A)_{i,j}$, for $(i,j) = (1,1), (1,2), (2,2)$, we again need $\tilde{A}^*_{21}$ to satisfy

$$e_2' \begin{bmatrix} (RA)_{11} & (RA)_{12} \\ \tilde{A}^*_{21} & (RA)_{22} \end{bmatrix}^{-1} = G^*_{2\cdot}R' \tag{1.B.25}$$

where the $(n-1)$ vector $\tilde{A}^*_{21}$ is the only unknown and the number of equations is $n(n-1)$. Hence, there are at most one $\tilde{A}^*_{21}$ to make (1.B.12) hold. □

Proof Theorem 1.6.

Proof. The proof is mostly the same as the proof in Theorem 1.2 - Theorem 1.4, because we can simply change the notations from the bivariate case to the multivariate case to complete most of the proof in Theorem 1.2 - Theorem 1.4. It is noteworthy that in the multivariate case, (1.25) and (1.26) now both represent $(n-1)$ equations, but the semiparametric Heckit estimator will still work for each one of these $(n-1)$ equations. □

Appendix 1.C Gibbs Sampling Scheme

1.C.1 Symmetry in Mixture of Normal Distributions

To easily identify all the parameters including the constant term, I impose symmetry in the mixture of normal distributions. In particular, for the $\bar{K}$ components, the first half of the components will have positive component mean, and the second half will have the same component variance but the opposite component mean compared to the first half. Technically speaking, we impose restrictions on μ_{ik} and σ_{ik}^2 so that

$$\begin{aligned} \mu_{i+\bar{K}/2,k} &= -\mu_{ik} \\ \sigma^2_{i+\bar{K}/2,k} &= \sigma^2_{ik} \\ \pi_{i+\bar{K}/2,k} &= \pi_{ik} \\ \forall i = 1, 2, \cdots, \bar{K}/2 & \end{aligned} \tag{1.C.1}$$

1.C.2 Prior

We impose conjugate priors on all the parameters. The details of the prior is stated below.[28]

Prior on A and A^*. Since the diagonal of A and A^* is normalized to be one, we can simply put independent prior for each off-diagonal element in A and A^*. For each (i, j) where $i \neq j$, the Gaussian prior is $A_{ij} \sim \mathcal{N}(0, 1)$ and $A^*_{ij} \sim \mathcal{N}(0, 1)$.

Prior on B. We can simply put independent prior for each element in B. For the lagged coefficient in equation i of variable j at lag l, the Gaussian prior is $B_{ijl} \sim \mathcal{N}(\mathbf{1}\{i = j\}, 2)$. For the constant term in B, the Gaussian prior is $B_{i0} \sim \mathcal{N}(0, 20^2)$.

Prior on μ_{ik}. We can simply put independent prior for each element μ_{ik} as $\mu_{ik} \sim \mathcal{N}(0, 1), k = 1, 2, \cdots, \bar{K}/2$.

Prior on σ^2_{ik}. We can simply put independent prior for each element σ^2_{ik} as $\sigma^2_{ik} \sim \text{IG}(5, 2), k = 1, 2, \cdots, \bar{K}/2$, where $\text{IG}(5, 2)$ means inverse-gamma with shape 5 and scale 2.

[28] The dataset is scaled up or down to make the vanilla VAR(p) model generate residual variance around 1. This will help all the following prior make sense.

PRIOR ON π_{ik}. For the i-th shock, the component probability needs to sum up to 1, i.e. $\sum_{k=1}^{\bar{K}} \pi_{ik} = 1$. Because we impose the symmetry, we need $\sum_{k=1}^{\bar{K}/2} 2\pi_{ik} = 1$, and thus we can simply put a Dirichlet prior on $(2\pi_{i1}, 2\pi_{i2}, \cdots, 2\pi_{i,\bar{K}/2})$ and get $(\pi_{i1}, \pi_{i2}, \cdots, \pi_{i,\bar{K}/2}) \propto \text{Dir}(1.5)$, where Dir(1.5) means Dirichlet prior with all the shape parameters equal to 1.5.

1.C.3 Gibbs Sampler

Now we denote all the parameters as the set $\theta = \{A, A^*, B, \pi_{ik}, \mu_{ik}, \sigma_{ik}^2\}$, and denote $\theta \backslash \theta_1$ as all the parameters except the subset θ_1.

With the augmented data, the likelihood can be evaluated by (1.40), and the posterior density with the augmented data will be

$$p(\theta, Z_{it}, y_{1t}^* | y_t) \propto p(\theta) \cdot p(Z_{it}|\theta) \cdot p(y_t, y_{1t}^* | \theta, Z_{it}) \tag{1.C.2}$$

We can now propose a Gibbs sampler to draw all the parameters θ, and the augmented data Z_{it}, y_{1t}^* alternatingly,[29]

- For $p(Z_{it}|\theta, y_{1t}^*, y_t)$:
  draw Z_{it} from the multinomial distribution, since $p(Z_{it} = k|\theta, y_{1t}^*, y_t) \propto \pi_k \phi(\varepsilon_{it}; \mu_{ik}, \sigma_{ik}^2)$
- For $p(\pi_{ik}|\theta \backslash \pi_{ik}, Z_{it}, y_{1t}^*, y_t)$:
  draw π_{ik} from the Dirichlet distribution, since $p(\pi_{ik}|\theta \backslash \pi_{ik}, Z_{it}, y_{1t}^*, y_t) \propto \pi_{ik}^{1.5-1+\sum_t \mathbf{1}\{Z_{it}=k\}}$
- For $p(B, \mu_{ik}|\theta \backslash \{B, \mu_{ik}\}, y_{1t}^*, y_t)$:
  set up dummy variables $\delta_{ikt} = \mathbf{1}\{Z_{it} = k\}$ for equation i, and then draw B and all the elements μ_{ik} jointly from the Gaussian posterior derived from WLS, where the weight for equation i at time t is $1/\sigma_{i,Z_{it}}$
- For $p(\sigma_{ik}^2|\theta \backslash \sigma_{ik}^2, y_{1t}^*, y_t)$:
  draw σ_{ik}^2 from the inverse-Gamma distribution, since
  $p(\sigma_{ik}^2|\theta \backslash \sigma_{ik}^2, y_{1t}^*, y_t) \propto (\sigma^2)^{-5-1-\frac{1}{2}\sum_t \mathbf{1}\{Z_{it}=k\}} \exp(-\frac{1}{\sigma^2}[2+0.5\sum_t (\varepsilon_{it} - \mu_{i,Z_{it}} \cdot \mathbf{1}\{Z_{it}=k\})^2])$
- For $p(A, A^*|\theta \backslash \{A, A^*\}, y_{1t}^*, y_t)$:
  draw A and A^* jointly using the Metropolis algorithm[30]
- For $p(y_{1t}^*|\theta, Z_{it}, y_t)$:
  draw y_{1t}^* from the truncated normal distribution, since we only need to consider the static correlation between y_{1t}^* and y_{2t} in the ZLB regime,

$$\begin{bmatrix} y_{1t}^* \\ y_{2t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_{1t}^* x_t + \mu_{1,Z_{1t}} \\ \beta_{2t}^* x_t + \mu_{2,Z_{2t}} \end{bmatrix}, G^* \begin{bmatrix} \sigma_{1,Z_{1t}}^2 & \\ & \sigma_{2,Z_{2t}}^2 \end{bmatrix} G^{*-1} \right)$$

[29] A detailed mathematical formula for the Gibbs sampling will be available in the online supplement appendix in the future version.

[30] To achieve better posterior sampling convergence, this paper draws each element of A and A^* individually conditional on other elements using the Metropolis algorithm, to explore the non-smooth posterior of A and A^*.

from which we can derive the truncated normal distribution of $y_{1t}^*|y_{1t}^* \leq 0, y_{2t}$

References

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, *24*(1-2), 3–61.

Aruoba, S. B., Cuba-Borda, P., Higa-Flores, K., Schorfheide, F., & Villalvazo, S. (2021). Piecewise-linear approximations and filtering for DSGE models with occasionally-binding constraints. *Review of Economic Dynamics*, *41*, 96–120.

Aruoba, S. B., Mlikota, M., Schorfheide, F., & Villalvazo, S. (2021). SVARs with occasionally-binding constraints. National Bureau of Economic Research.

Bernanke, B. S. (2020). The new tools of monetary policy. *American Economic Review*, *110*(4), 943–83.

Braun, R. (2021). The importance of supply and demand for oil prices: evidence from non-Gaussianity. Unpublished manuscript.

Brunnermeier, M., Palia, D., Sastry, K. A., & Sims, C. A. (2021). Feedbacks: Financial markets and economic activity. *American Economic Review*, *111*(6), 1845–79.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, *32*(2), 189–218.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, *36*(3), 287–314.

Cosslett, S. R. (1984). Distribution-free estimator for the sample selection model. Unpublished manuscript.

Debortoli, D., Galí, J., & Gambetti, L. (2020). On the empirical (ir)relevance of the zero lower bound constraint. *NBER Macroeconomics Annual*, *34*(1), 141–170.

Eggertsson, G. B. & Woodford, M. (2003). Optimal monetary policy in a liquidity trap. National Bureau of Economic Research.

Fiorentini, G. & Sentana, E. (2020). Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. CEMFI.

Gertler, M. & Karadi, P. (2011). A model of unconventional monetary policy. *Journal of Monetary Economics*, *58*(1), 17–34.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gouriéroux, C., Laffont, J.-J., & Monfort, A. (1980). Coherency conditions in simultaneous linear equation models with endogenous regimes. *Econometrica*, *48*(3), 675–696.

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

Gunsilius, F. & Schennach, S. (2021). Independent nonlinear component analysis. *Journal of the American Statistical Association*, *0*(0), 1–14.

Honore, B. E., Kyriazidou, E., & Udry, C. (1997). Estimation of type 3 tobit models using symmetric trimming and pairwise comparisons. *Journal of Econometrics*, *76*(1-2), 107–128.

Hyvarinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. Wiley.

Jarocinski, M. (2021). Estimating Fed's unconventional policy shocks. ECB Working Paper.

Kuttner, K. N. (2018). Outside the box: Unconventional monetary policy in the great recession and beyond. *Journal of Economic Perspectives*, *32*(4), 121–46.

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

Lin, J. K. (1998). Factorizing multivariate function classes. *Advances in Neural Information Processing Systems*, *0*, 563–569.

Mavroeidis, S. (2021). Identification at the Zero Lower Bound. *Econometrica*, *forthcoming*.

Newey, W. K., Powell, J. L., & Walker, J. R. (1990). Semiparametric estimation of selection models: Some empirical results. *American Economic Review*, *80*(2), 324–328.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, *25*(3), 303–325.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of Macroeconomics*, *2*, 71–162.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, *56*(4), 931–954.

Sims, C. A. (2020). SVAR identification through heteroskedasticity with misspecified regimes. Unpublished manuscript.

Sims, C. A. & Zha, T. (2006). Were there regime switches in US monetary policy? *American Economic Review*, *96*(1), 54–81.

Sims, E. R. & Wu, J. C. (2019). The four equation new keynesian model. National Bureau of Economic Research.

Chapter 2

Local Projections vs. VARs: Lessons From Thousands of DGPs

2.1 Introduction

Since Jordà (2005) introduced the popular local projection (LP) impulse response estimator, there has been a debate about its benefits and drawbacks relative to Vector Autoregression (VAR) estimation (Sims, 1980).[1] Recently, Plagborg-Møller & Wolf (2021) proved that these two methods in fact estimate precisely the same impulse responses asymptotically, provided that the lag length used for estimation tends to infinity. This result holds regardless of identification scheme and regardless of the underlying data generating process (DGP). Nevertheless, the question of which estimator to choose in finite samples remains open. It is also an urgent question, since researchers have remarked that LPs and VARs can give conflicting results when applied to central economic questions such as the effects of monetary or fiscal stimulus (e.g., Ramey, 2016; Nakamura & Steinsson, 2018).

Whereas the LP estimator utilizes the sample autocovariances flexibly by directly projecting an outcome at the future horizon h on current covariates, a VAR(p) estimator instead extrapolates longer-run impulse responses from the first p sample autocovariances. Hence, though the estimates from the two methods agree approximately at horizons $h \leq p$, they can disagree substantially at intermediate and long horizons.[2] Intuitively, the extrapolation employed by VARs should yield a lower variance but potentially a higher bias than for LPs, perfectly analogous to the trade-off between direct and iterated reduced-form forecasts (Schorfheide, 2005; Kilian & Lütkepohl, 2017).[3] How much more should one care about bias than variance

[1]Author: Dake Li dakel@princeton.edu, Mikkel Plagborg-Møller mikkelpm@princeton.edu, and Christian K. Wolf ckwolf@mit.edu. We received helpful comments from Isaiah Andrews, Régis Barnichon, Gabe Chodorow-Reich, Viet Hoang Dinh, Òscar Jordà, Helmut Lütkepohl, Massimiliano Marcellino, Pepe Montiel Olea, Ulrich Müller, Emi Nakamura, Chris Sims, Lumi Stevens, Jim Stock, Mark Watson, Andrei Zeleneev, and numerous seminar and conference participants. Plagborg-Møller acknowledges that this material is based upon work supported by the NSF under Grant #1851665. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

[2]See Plagborg-Møller & Wolf (2021, Proposition 2) for a formal result.

[3]The trade-off is also conceptually similar to the relationship between polynomial series estimators and kernel estimators in cross-sectional nonparametric regression.

to optimally choose the LP estimator over the VAR estimator in realistic sample sizes? And how does the trade-off depend on the DGP? Unfortunately, these questions are challenging to answer analytically, due to the dynamic and nonlinear nature of the time series estimators, as well as the breadth of DGPs encountered in applied practice.

In this paper we illuminate the bias-variance trade-off in impulse response estimation through a comprehensive simulation study, applying LP and VAR methods to thousands of empirically relevant DGPs. Our goal is to identify which estimators perform well *on average* across many DGPs and thus may serve as practical default procedures. Rather than insisting on the usual binary distinction between "local projections" and "VARs", we furthermore consider an entire menu of related estimation approaches. We find that the usual least-squares LP estimator tends to have lower bias than the least-squares VAR estimator, as expected, but this bias reduction incurs a substantial cost of higher variance. Consequently, unless researchers are almost exclusively concerned with bias, least-squares LP is not optimal, and shrinkage estimation via Bayesian VARs or penalized LPs is usually attractive.

Our simulation study considers an extensive array of DGPs, obtained by drawing specifications at random from a large-scale, empirically calibrated dynamic factor model (DFM).[4] We use the estimated DFM in Stock & Watson (2016), which has been fitted to 207 U.S. macroeconomic time series that span a wide variety of variable categories. As discussed by Stock & Watson, DFMs are well-known to accurately capture the joint co-movements of conventional macroeconomic data, thus ensuring that our simulation results will be informative about the universe of standard U.S. time series. From the encompassing 207-variable DFM we draw 6,000 random subsets of five variables (subject to certain constraints that emulate applied practice), whose model-implied time series processes constitute the set of DGPs we consider for our simulation study. Importantly, these DGPs vary in terms of how well they can be approximated by low-order VAR models, thus yielding a non-trivial bias-variance trade-off between LPs and VARs. Moreover, the DGPs exhibit substantial heterogeneity in persistence, shape of impulse response functions, and invertibility of the structural shocks, consistent with the heterogeneity faced by applied researchers. An estimation method that works well across this multitude of DGPs therefore carries substantial promise as an attractive default procedure.

We study the ability of several variants of LP and VAR methods to accurately estimate impulse response functions. In addition to the popular least-squares LP and VAR estimators, we enrich the bias-variance possibility frontier by considering: (i) penalized LP (Barnichon & Brownlees, 2019), which smooths out the estimated impulse response function; (ii) Bayesian VAR estimation, which shrinks VAR coefficients toward white noise; (iii) model averaging of univariate and multivariate VAR models of various lag lengths (Hansen, 2016); and (iv) bias correction of the VAR coefficients (Pope, 1990; Kilian, 1998). For each estimation method, we consider three classical structural identification schemes: observed shocks, instrumental variables (IVs)/proxies, and recursive identification. For IV identification, we distinguish between internal IV methods (Ramey, 2011; Plagborg-Møller & Wolf, 2021) and external IV methods (Stock, 2008; Stock & Watson, 2012; Mertens & Ravn, 2013).

Applying the various estimation methods to simulated data from our multitude of DGPs, a clear and unavoidable bias-variance trade-off emerges. We highlight four main lessons:

[4] Our overall approach is inspired by Lazarus et al. (2018), who are instead interested in the question of how to select among different long-run variance estimators.

1. Least-squares LP and VAR estimators lie on opposite ends of the bias-variance spectrum: small bias and large variance for LPs, and large bias and small variance for VARs. Thus, for any given impulse response horizon, there exists a weight on squared bias relative to variance in the loss function that would make a researcher indifferent between the least-squares LP and VAR estimators. Yet, the *slope* of this trade-off is stark: indifference between the two methods requires a researcher to care on average around four times more about squared bias than about variance.

2. Shrinkage methods dramatically lower the variance of LP and VAR methods, at a usually moderate cost in terms of bias. In particular, unless researchers care almost exclusively about bias, penalized LP is preferred to least-squares LP, and Bayesian VAR shrinkage is preferred to or at least competitive with the least-squares VAR estimator. However, when the data is very persistent, the performance of the Bayesian VAR estimator is sensitive to the choice of prior at long horizons.

3. For any given loss function, no single method dominates at all horizons. Unless the concern for bias is overwhelming, penalized LP is the most attractive estimation method at short horizons, while least-squares or Bayesian VAR estimation are the most attractive methods at intermediate and long horizons.

4. In the case of IV identification, the SVAR-IV estimator is heavily (median-)biased, but provides substantial reduction in dispersion. Depending on the weight attached to bias, it may therefore be justifiable to use external IV methods despite their lack of robustness to non-invertibility (unlike internal IV methods).

Our findings provide a novel perspective on recent work emphasizing the potential dangers of VAR model mis-specification (Ramey, 2016; Nakamura & Steinsson, 2018). We consider DGPs that do not admit finite-order VAR representations, so VAR methods indeed suffer from larger bias, as cautioned there. Reducing that bias via direct projection, however, incurs a steep cost in terms of increased sampling variance at intermediate and long horizons. Researchers who employ conventional LP estimators should be prepared to pay that price.

Literature. Our large-scale simulation study is inspired by the seminal work of Marcellino et al. (2006) on direct and iterated multi-step forecasts, though we focus instead on structural impulse responses. While simulation studies in the forecasting literature often consider low-dimensional specifications, we consider systems with several variables, consistent with standard practice in the applied structural macroeconometrics literature. The structural perspective also requires us to contend with issues such as the variety of different popular shock identification schemes, normalization of impulse responses, and the special role of external instrumental variables.

Our large-scale set-up differs from prior simulation studies of LP and VAR methods, which have considered at most a handful of DGPs. Examples include Jordà (2005), Meier (2005), Kilian & Kim (2011), Brugnolini (2018), Choi & Chudik (2019), Austin (2020), and Bruns & Lütkepohl (2021). These papers either obtain their DGPs from stylized, low-dimensional VARMA models, calibrated DSGE models, and/or a few empirically calibrated VAR models. Our analysis also differs in the following respects: we consider shrinkage estimation procedures as competitors to the least-squares estimators; we study several popular structural identification schemes; and we examine how our conclusions vary with the impulse response horizon and the researcher's loss function. All these features are essential to the above-mentioned main lessons that we draw from our results.

Even though the simulation results are at the heart of our analysis, we start off by illustrating the bias-variance trade-off through an analytical example that builds on Schorfheide (2005). That paper develops a general theory of the asymptotic bias and variance of direct and iterated (reduced-form) forecasts under local mis-specification. While these theoretical results are valuable for analytically distilling the forces at work, they do not by themselves resolve the bias-variance trade-off faced by practitioners, as this trade-off invariably depends in a complicated fashion on many features of the DGP.

Finally, we stress that our paper focuses solely on point estimation, as opposed to inference or hypothesis testing. See Inoue & Kilian (2020) and Montiel Olea & Plagborg-Møller (2021) for theoretical and simulation results on VAR and LP confidence interval procedures. Moreover, we focus exclusively on impulse response estimands, rather than variance decompositions or historical decompositions.

Outline. Section 2.2 illustrates the bias-variance trade-off for LP and VAR estimators using a simple analytical example. Section 2.3 describes the empirically calibrated dynamic factor model that we use to generate our many DGPs. Section 2.4 defines the menu of LP- and VAR-based estimation procedures. Section 2.5 contains our main simulation results and robustness checks. Section 2.6 summarizes the lessons for applied researchers and offers guidance for future research. The appendix contains implementation details. A supplemental appendix with proofs and further simulation results and a Matlab code suite are available online (`https://github.com/dake-li/lp_var_simul`).

2.2 The bias-variance trade-off

This section motivates our simulation study with an analytical discussion of the bias-variance trade-off between LP and VAR impulse response estimators. Section 2.2.1 begins with a simple model that cleanly illustrates the trade-off, and Section 2.2.2 connects the discussion to the rest of the paper.

2.2.1 Illustrative example

Plagborg-Møller & Wolf (2021) show that the impulse response estimands of VAR and LP estimators with p lags generally differ at horizons $h > p$: the VAR extrapolates from the first p autocovariances of the data, while LP exploits all autocovariances out to horizon $h + p$. This observation suggests the presence of a bias-variance trade-off whenever the true DGP is not a finite-order VAR, perfectly analogous to the choice between "direct" and "iterated" predictions in multi-step forecasting (Marcellino et al., 2006). We here formalize this basic intuition by extending the arguments of Schorfheide (2005) to structural impulse response estimation in a simple DGP.

Model. Consider a simple sequence of drifting DGPs:

$$y_t = \rho y_{t-1} + \varepsilon_{1,t} + \varepsilon_{2,t} + \frac{\alpha}{\sqrt{T}}\varepsilon_{2,t-1}, \tag{2.1}$$

where $\varepsilon_t \equiv (\varepsilon_{1,t}, \varepsilon_{2,t})'$ is an i.i.d. white noise process with $\mathrm{Var}(\varepsilon_t) = \mathrm{diag}(1, \sigma_2^2)$. The DGP drifts towards an AR(1) process at rate $T^{-1/2}$, where T is the sample size. We will show that this ensures a non-trivial bias-variance trade-off in the limit $T \to \infty$. The DGP captures the notion that finite-order autoregressive

models are often a good—but not exact—approximation to the true underlying DGP. Note that the degree of autoregressive mis-specification is governed by α.

We are interested in the impulse responses of y_t with respect to a unit impulse in $\varepsilon_{1,t}$. The true impulse response function is evidently $\theta_h \equiv \rho^h$. We assume that the researcher observes $w_t \equiv (\varepsilon_{1,t}, y_t)'$, i.e., she observes the shock $\varepsilon_{1,t}$ but not $\varepsilon_{2,t}$. To evaluate the performance of a given estimator $\hat{\theta}_h$, we will throughout this paper consider loss functions of the form

$$\mathcal{L}_\omega(\theta_h, \hat{\theta}_h) = \omega \times \Big(\mathbb{E}[\hat{\theta}_h - \theta_h]\Big)^2 + (1-\omega) \times \mathrm{Var}(\hat{\theta}_h). \tag{2.2}$$

For $\omega = \frac{1}{2}$, this is proportional to the mean squared error (MSE).[5] For $\omega > \frac{1}{2}$, the researcher is more concerned about (squared) bias than variance, and *vice versa*.

ESTIMATORS. For now, we consider two estimators of θ_h.

1. **LP.** The least-squares local projection estimator $\hat{\beta}_h$ is obtained from the OLS regression

$$y_{t+h} = \hat{\beta}_h \varepsilon_{1,t} + \hat{\zeta}_h' w_{t-1} + \mathrm{residual}_{t,h}, \tag{2.3}$$

   at each horizon h. Notice that this LP specification controls for one lag of the data.

2. **VAR.** We consider a recursive VAR specification in $w_t = (\varepsilon_{1,t}, y_t)'$, again with one lag. Define the usual least-squares coefficient estimator $\hat{A} \equiv (\sum_{t=2}^T w_t w_{t-1}')(\sum_{t=2}^T w_{t-1} w_{t-1})^{-1}$ and residual covariance matrix $\hat{\Sigma} \equiv T^{-1}\sum_{t=2}^T \hat{u}_t \hat{u}_t'$, where $\hat{u}_t \equiv w_t - \hat{A} w_{t-1}$. Define the lower triangular Cholesky factor $\hat{C}$, where $\hat{C}\hat{C}' = \hat{\Sigma}$. The un-normalized VAR impulse responses with respect to the first orthogonalized shock at horizon h are given by $\hat{A}^h \hat{C} e_1$, where e_j is the j-th unit vector of dimension 2, $j = 1, 2$. To facilitate comparison with LP, we normalize the impact response of the first variable in the VAR (i.e., $\varepsilon_{1,t}$) with respect to the first shock to be 1. This yields the estimator $\hat{\delta}_h \equiv e_2' \hat{A}^h \hat{\gamma}$, where $\hat{\gamma} \equiv (1, \hat{\kappa})'$ and $\hat{\kappa} \equiv \hat{\Sigma}_{21}/\hat{\Sigma}_{11}$.[6]

Note that at the impact horizon $h = 0$, the two estimators $\hat{\beta}_0$ and $\hat{\delta}_0$ are numerically equal.

TRADE-OFF. How does the optimal choice of estimation method depend on the properties of the DGP (2.1) and the bias weight ω in the loss function (2.2)? Along the stated asymptote, the researcher faces a clear bias-variance trade-off:

Proposition 2.1. *Consider the model (2.1), and fix $h \geq 0$, $\rho \in (-1, 1)$, $\sigma_2 > 0$, and $\alpha \in \mathbb{R}$. Assume $E(\varepsilon_{j,t}^4) < \infty$ for $j = 1, 2$. Define $\sigma_{0,y}^2 \equiv \frac{1+\sigma_2^2}{1-\rho^2}$. Then, as $T \to \infty$,*

$$\sqrt{T}(\hat{\beta}_h - \theta_h) \xrightarrow{d} N(\mathrm{aBias}_{LP}, \mathrm{aVar}_{LP}), \quad \sqrt{T}(\hat{\delta}_h - \theta_h) \xrightarrow{d} N(\mathrm{aBias}_{VAR}, \mathrm{aVar}_{VAR}), \tag{2.4}$$

[5] The objective function (2.2) is not a loss function in the usual decision theoretic sense (which would call it a risk function when $\omega = \frac{1}{2}$). We proceed with the non-standard terminology for ease of exposition.

[6] We have $\hat{C} = \begin{pmatrix} \sqrt{\hat{\Sigma}_{11}} & 0 \\ \hat{\Sigma}_{21}/\sqrt{\hat{\Sigma}_{11}} & \hat{\Sigma}_{22} - \hat{\Sigma}_{21}^2/\hat{\Sigma}_{11} \end{pmatrix}$. We therefore achieve the desired normalization of the impact effect of the shock by dividing $\hat{C}e_1$ by $\sqrt{\hat{\Sigma}_{11}}$. This gives the normalized impulse responses $\hat{A}^h\hat{\gamma}$.

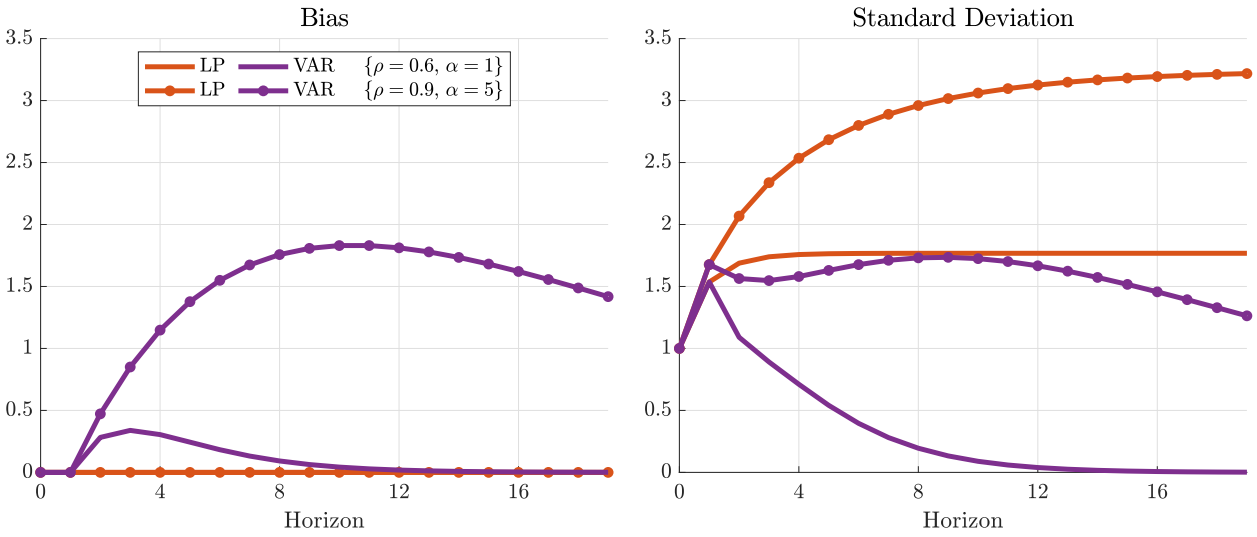


Figure 2.1: Asymptotic bias and standard deviation for LP (red) and VAR (purple) in the DGP (2.1) with $\sigma_1 = 1$ and $\{\rho = 0.6, \alpha = 1\}$ (no markers) or $\{\rho = 0.9, \alpha = 5\}$ (markers).

where for all $h \geq 0$,

$$\text{aBias}_{LP} \equiv 0, \quad \text{aVar}_{LP} \equiv \sigma^2_{0,y}(1 - \rho^{2(h+1)}) - \rho^{2h},$$

and for $h \geq 1$,

$$\text{aBias}_{VAR} \equiv \rho^{h-1}(h-1)\frac{\alpha\sigma_2^2}{\sigma^2_{0,y} - 1}, \quad \text{aVar}_{VAR} \equiv \rho^{2(h-1)}(1-\rho^2)\sigma^2_{0,y}\left(1 + \frac{(h-1)^2}{\sigma^2_{0,y} - 1}\right) + \rho^{2h}\sigma_2^2.$$

Proof. Please see Supplemental Appendix F. □

Proposition 2.1 implies a ranking of the asymptotic biases and variances. First, we clearly have $|\,\text{aBias}_{\text{VAR}}\,| > |\,\text{aBias}_{\text{LP}}\,| = 0$ whenever $h \geq 2$, $\rho \neq 0$, and $\alpha \neq 0$. Second, as in Schorfheide (2005), the asymptotic variances in Proposition 2.1 are the same as in the well-specified model with $\alpha = 0$. It then follows from the Cramér-Rao bound that $\text{aVar}_{\text{VAR}} \leq \text{aVar}_{\text{LP}}$ for all h, and the inequality is strict when $h \geq 2$ and $\rho \neq 0$.[7]

Figure 2.1 plots the asymptotic bias and standard deviation for two parametrizations: one with low persistence and modest mis-specification ($\rho = 0.6$ and $\alpha = 1$), and one with high persistence and severe mis-specification ($\rho = 0.9$ and $\alpha = 5$).[8] Both set $\sigma_2 = 1$. Figure 2.2 then plots the weight $\omega = \omega_h^*$ on

[7]The argument goes as follows. Both asymptotic variances in Proposition 2.1 are the same as they would be when estimating the parameter $\tilde{\theta}_h \equiv e_2' A^h \text{chol}(\Sigma) e_1 / \Sigma_{11}^{1/2}$ in the well-specified bivariate VAR(1) model $w_t = Aw_{t-1} + u_t$, $u_t \overset{i.i.d.}{\sim} N(0, \Sigma)$, when $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & \rho \end{smallmatrix}\right)$ and $\Sigma = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1+\sigma_2^2 \end{smallmatrix}\right)$ ("chol" denotes the lower triangular Cholesky factor). The VAR estimator is the MLE in this model. It can be verified that the LP estimator (which regresses $w_{2,t+h}$ on $w_{1,t}$, controlling for w_{t-1}) is a consistent and asymptotically regular estimator of $\tilde{\theta}_h$ in this model (regardless of A and Σ). Hence, the Cramér-Rao bound implies $\text{aVar}_{\text{VAR}} \leq \text{aVar}_{\text{LP}}$. Moreover, when $h \geq 2$ and $\rho \neq 0$, LP is strictly inefficient since its asymptotic correlation with the VAR estimator is not 1, according to the proof of Proposition 2.1.

[8]To interpret the magnitude of α, let $\hat{\tau}$ denote the t-statistic for testing the significance of the second lag in a univariate AR(2) regression for $\{y_t\}$. Supplemental Appendix F.3 shows that $\hat{\tau} \xrightarrow{d} N\left(-\rho\frac{\sigma_2^2}{1+\sigma_2^2}\alpha, 1\right)$. We remind the reader, however,

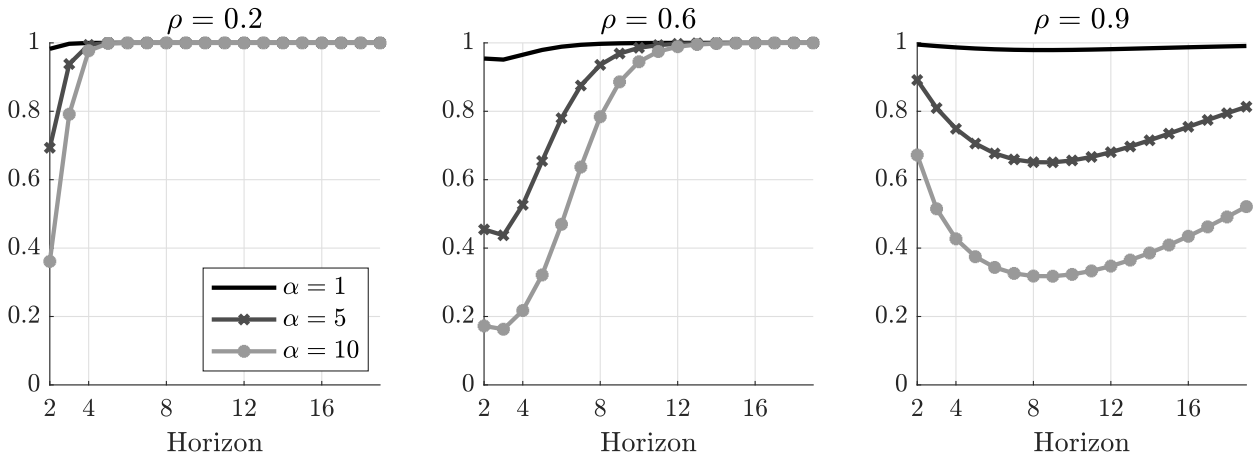


Figure 2.2: Weight $\omega = \omega_h^*$ in the asymptotic loss function $\omega \times \text{aBias}^2_{\hat{\theta}_h} + (1-\omega) \times \text{aVar}_{\hat{\theta}_h}$ that yields indifference between the LP and VAR estimator. LP is preferred whenever $\omega \geq \omega_h^*$. The three panels correspond to different values of $\rho \in \{0.2, 0.6, 0.9\}$; the three curves in each panel correspond to different values of $\alpha \in \{1, 5, 10\}$. All results are computed with $\sigma_2 = 1$. The figure omits the horizons $h \in \{0, 1\}$, at which the two estimation methods are (asymptotically) equivalent.

bias in the asymptotic analogue of the loss function (2.2) that yields indifference between the LP and VAR estimator. That is, LP is preferred over VAR if and only if $\omega \geq \omega_h^*$. The three panels correspond to different degrees of persistence $\rho \in \{0.2, 0.6, 0.9\}$, while the three curves in each panel correspond to different degrees of mis-specification $\alpha \in \{1, 5, 10\}$.

We draw the following four conclusions from Proposition 2.1 and Figures 2.1 and 2.2:

1. For $h \in \{0, 1\}$, there actually is no bias-variance trade-off: on impact, the two estimators are numerically equivalent; at $h = 1$, both are asymptotically unbiased, and the asymptotic variance coincides. Thus, in Figure 2.1, the red and purple lines coincide initially, for both parametrizations. Intuitively, the equivalence at $h = 1$ reflects the fact that the VAR(1) estimator does not extrapolate, instead reporting the direct projection of y_{t+1} on w_t, exactly as LP does (Plagborg-Møller & Wolf, 2021).

2. For $h \geq 2$, the bias-variance trade-off is non-trivial. LP directly projects y_{t+h} on the shock $\varepsilon_{1,t}$, which is uncorrelated with any lagged controls, so the asymptotic bias is always zero. In contrast, the VAR(1) estimator extrapolates the response at horizon h from the sample autocovariances at horizon 1. Though this tight parametric extrapolation yields a low variance relative to LP, it incurs a bias due to dynamic mis-specification when $\alpha \neq 0$. It follows that the indifference weight ω_h^* in Figure 2.2 is always in $(0, 1)$: LP can be justified if the concern for bias is sufficiently high, and *vice versa* for VAR.

3. The indifference weight ω_h^* is decreasing in the degree of mis-specification α. That is, the greater the mis-specification, the less the researcher needs to care about bias to prefer LPs over VARs. This result follows immediately from the observation that the asymptotic variances are independent of α, while the

that the purpose of this section is illustration, not quantitative realism.

asymptotic bias of the VAR is increasing in α.[9]

4. For $h \to \infty$, the bias-variance trade-off is invariably resolved in favor of the VAR estimator: both asymptotic bias and variance for the VAR tend to zero, while the asymptotic variance of LP tends to $\sigma^2_{0,y} > 0$. The practical relevance of this result hinges crucially on the persistence of the DGP, since Figure 2.2 shows that the indifference weight ω^*_h converges more slowly to 1 as $h \to \infty$ when the DGP is more persistent. The figure also shows that ω^*_h need not be a monotone function of the horizon h.

To summarize, even in this simple model environment, the researcher's evaluation of VAR and LP estimators is shaped by several parameters: the DGP's persistence ρ, the degree of mis-specification α, the horizon h of interest, and the weight ω attached to squared bias rather than variance in the loss function.

2.2.2 Outlook

Because analytical bias-variance calculations invariably end up depending on a multitude of parameters, the rest of this paper will use simulations to explore the trade-off across a rich and empirically relevant set of DGPs. In the language of Section 2.2.1, these DGPs will tell us about empirically plausible degrees of mis-specification α and persistence ρ, and so about the practically relevant bias weight ω necessary to justify the use of one linear projection technique over another. Moreover, we will also consider several variants of the least-squares LP and VAR estimators that use shrinkage approaches to further trace out the bias-variance possibility frontier.

2.3 Data generating processes

This section presents our DGPs. We define the empirically calibrated encompassing model in Section 2.3.1, from which we then draw thousands of DGPs with corresponding structural impulse response estimands, as described in Section 2.3.2. We discuss implementation details in Section 2.3.3, and provide summary statistics for the DGPs in Section 2.3.4. Various modifications to this baseline set of DGPs are considered later in Section 2.5.5.

2.3.1 Encompassing model

We construct our simulation DGPs from an encompassing model that is known to accurately describe the population of U.S. macroeconomic time series: the large-scale dynamic factor model (DFM) of Stock & Watson (2016).

The DFM postulates that a large-dimensional $n_X \times 1$ vector X_t of observed macroeconomic time series is driven by a low-dimensional $n_f \times 1$ vector f_t of latent factors, as well as an $n_X \times 1$ vector v_t of idiosyncratic components. The latent factors are assumed to follow a stationary VAR(p) process

$$f_t = \Phi(L) f_{t-1} + H\varepsilon_t, \tag{2.5}$$

[9] Schorfheide (2005), shows that this result goes through in much more general DGPs, though he focuses on reduced-form estimators.

where $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n_f,t})'$ is an $n_f \times 1$ vector of aggregate shocks, which are i.i.d. and mutually uncorrelated, $\text{Var}(\varepsilon_t) = I_{n_f}$. The $n_f \times n_f$ matrix H determines the impact impulse responses of the factors with respect to the aggregate shocks. The observed macroeconomic aggregates X_t are given by

$$X_t = \Lambda f_t + v_t, \tag{2.6}$$

where the idiosyncratic component $v_{i,t}$ for macro observable $X_{i,t}$ follows the AR(q) process

$$v_{i,t} = \Delta_i(L)v_{i,t-1} + \Xi_i \xi_{i,t}, \tag{2.7}$$

with $\xi_{i,t}$ i.i.d. across t and i. We assume all shocks and innovations are jointly normal and homoskedastic. In Section 2.3.3 we describe how the parameters of the DFM are calibrated to the Stock & Watson (2016) data set, but first we describe how we construct our lower-dimensional DGPs from the encompassing large-scale DFM.

2.3.2 DGPs and impulse response estimands

We use the encompassing model (2.5)–(2.7) to build thousands of lower-dimensional DGPs for our simulation study. Specifically, for each DGP, we draw a random subset of $n_{\bar{w}}$ variables $\bar{w}_t$ from the large vector X_t, i.e., $\bar{w}_t \subset X_t$. The variables $\bar{w}_t$ follow the time series process implied by the encompassing model (2.5)–(2.7). In particular, $\bar{w}_t$ is driven by some combination of aggregate structural shocks ε_t and idiosyncratic components v_t. We draw thousands of such random combinations of variables, thus yielding thousands of lower-dimensional DGPs. The details of how we select the variable combinations are postponed until Section 2.3.3.

For each DGP drawn in this way, we consider three types of structural impulse response estimands, chosen to mimic as closely as possible popular structural identification schemes in applied macroeconometrics (Ramey, 2016; Stock & Watson, 2016). In the following, $y_t \in \bar{w}_t$ denotes a response variable of interest in the DGP, $i_t \in \bar{w}_t$ is a variable used to normalize the scale of the shock (to be defined below), z_t is an external instrument (if applicable), and w_t denotes the vector of all observed time series in the DGP.

1. **Observed shock identification.** In this identification scheme we assume that the econometrician observes both the endogenous variables $\bar{w}_t$ and the first structural shock $\varepsilon_{1,t}$, so the full vector of observables is $w_t = (\varepsilon_{1,t}, \bar{w}_t)'$. The objects of interest are the impulse responses of an outcome variable y_t with respect to a one standard deviation (i.e., one unit) innovation to $\varepsilon_{1,t}$:

$$\theta_h \equiv \bar{\Lambda}_{\iota_y,\bullet}\Theta^f_{\bullet,1,h}, \quad h = 0, 1, 2, \ldots, \tag{2.8}$$

   where $\Theta^f(L) = (I - \Phi(L))^{-1}H$ are the impulse responses of the factors f_t to the structural shocks ε_t implied by (2.5), while $\bar{\Lambda}$ are those rows of Λ that correspond to the observables $\bar{w}_t$. The index ι_y corresponds to the location of y_t in the vector $\bar{w}_t$.

   This set-up captures those empirical studies in which the researcher has constructed a plausible direct measure of the shock of interest. Examples include the monetary shock series of Romer & Romer (2004) or the fiscal shock series of Ramey (2011).

2. **IV/proxy identification.** In this scheme, rather than directly observing the structural shock $\varepsilon_{1,t}$, the econometrician observes the noisy proxy

$$z_t = \rho_z z_{t-1} + \varepsilon_{1,t} + \nu_t, \tag{2.9}$$

where ν_t is an i.i.d. process (independent of all shocks and innovations in the DFM) with $\text{Var}(\nu_t) = \sigma_\nu^2$. The full vector of observables is thus $w_t = (z_t, \bar{w}_t)'$. As is standard in IV applications, we here adopt the "unit effect" normalization of Stock & Watson (2016), so the object of interest becomes

$$\theta_h \equiv \frac{\bar{\Lambda}_{\iota_y,\bullet}\Theta^f_{\bullet,1,h}}{\bar{\Lambda}_{\iota_i,\bullet}\Theta^f_{\bullet,1,0}}, \quad h = 0, 1, 2, \ldots, \tag{2.10}$$

where the index ι_i corresponds to the location of some normalization variable i_t in the vector $\bar{w}_t$. The above unit effect normalization defines the magnitude of the shock $\varepsilon_{1,t}$ such that it raises the normalization variable i_t by one unit on impact.

One example of an IV z_t is the high-frequency change in futures prices around monetary policy announcements employed by Gertler & Karadi (2015) to identify the effects of monetary policy shocks.

3. **Recursive identification.** For our third and final structural estimand, the researcher observes only the endogenous variables $\bar{w}_t \subset X_t$, with no further direct or noisy shock measures. Thus, the total vector of observables is $w_t = \bar{w}_t$. Consistent with a large literature on recursive shock identification in VARs (e.g., Christiano et al., 1999; Blanchard & Perotti, 2002), we take as the estimand the impulse responses with respect to a recursive orthogonalization of the reduced-form (Wold) forecast errors in the VAR(∞) process for $\bar{w}_t$ implied by the DFM. The ordering of the variables for the recursive identification is described below in Section 2.3.3. Note that the recursively orthogonalized innovation differs across DGPs, and it generally does not equal any of the structural shocks $\varepsilon_{j,t}$ in the DFM. We nevertheless consider this impulse response estimand due to its popularity in applied work. We relegate the mathematical definition of the recursive impulse response estimand to Supplemental Appendix C.

2.3.3 Implementation

This section first discusses the empirical calibration of the DFM and then specifies the particular DGPs and structural impulse responses that we consider in the simulation study.

DFM PARAMETERS. We parametrize the DFM (2.5)–(2.7) based on the empirical reduced-form parameter estimates from Stock & Watson (2016), using the same specification as in Lazarus et al. (2018). We provide a brief summary here and refer to Stock & Watson for further details. The full vector of observables X_t contains quarterly observations on 207 time series for 1959Q1–2014Q4, mostly consisting of real activity variables, price measures, interest rates, asset and wealth variables, and productivity series.[10] Each series is seasonally adjusted and—importantly—transformed to approximate stationarity. Following the formal dimensionality tests of Stock & Watson, we allow for $n_f = 6$ factors and two lags in the factor equation (2.5). Finally, we

[10] Table 1 and the Data Appendix of Stock & Watson (2016) list all variables and their categories.

allow for two lags in the idiosyncratic component equation (2.7). The reduced-form parameters are estimated by principal components and least-squares procedures. This pins down all parameters of the DFM except for the structural impact response matrix H, which we discuss below.

DGP AND ESTIMAND SELECTION. To provide a comprehensive picture of the bias-variance trade-off, we select thousands of different possible sets of observables $\bar{w}_t \subset X_t$. We consider two protocols for selecting these observables: one aimed at mimicking monetary policy shock applications, and one aimed at fiscal policy shock applications. Specifically, for each type of policy shock, we randomly draw $n_{\text{spec}} = 3{,}000$ configurations of $n_{\bar{w}} = 5$ macro observables $\bar{w}_t$. Thus, we end up with a total of 6,000 DGPs. For the monetary policy DGPs we restrict $\bar{w}_t$ to always contain the federal funds rate, while for the fiscal policy DGPs we restrict $\bar{w}_t$ to contain federal government spending. These two series are chosen as the normalization variables i_t for the IV and recursive estimands. The remaining four variables in $\bar{w}_t$ are selected uniformly at random from X_t, except we impose that at least one variable should be a measure of real activity, and at least one other variable a measure of prices.[11] The impulse response variable y_t is selected uniformly at random from the four series (other than i_t).

For each of the 6,000 DGPs, we implement the three structural impulse response estimands in Section 2.3.2 as follows:

1. **Observed shock.** We select the structural impact response matrix H in the factor equation (2.5) so as to maximize the impact effect of the shock $\varepsilon_{1,t}$ on the federal funds rate (for monetary shocks) and government spending (for fiscal shocks), subject to the constraint that H is consistent with our estimate of the reduced-form innovation variance-covariance matrix for the factors. This ensures that monetary and fiscal shocks account for substantial short-run variation in nominal interest rates and government spending, respectively. Additionally, we avoid issues related to division by near-zeros when normalizing the impulse responses for the IV estimand. See Section 2.A.1 for further details.

2. **IV.** The matrix H is defined as in the "observed shock" case. As for the IV parameters in equation (2.9), we draw ρ_z uniformly at random from the set $\{0, 0.25, 0.5\}$.[12] To ensure an empirically plausible signal-to-noise ratio, we calibrate σ_ν^2 using real-world IV series. Specifically, we base the calibration on the Romer & Romer (2004) series for monetary policy DGPs and the Ben Zeev & Pappa (2017) series for fiscal policy DGPs. See Section 2.A.2 for details. We show below in Section 2.3.4 that the strength of these IVs ranges from somewhat weak to moderately strong, as measured by the first-stage F-statistic.

3. **Recursive.** For monetary policy DGPs, we order the federal funds rate last, as in Christiano et al. (1999); this restricts the other included variables to not respond contemporaneously to the monetary innovation. For fiscal policy DGPs, we order the government expenditure series first, as in Blanchard & Perotti (2002); this restricts the fiscal authority to respond to other innovations in the recursive VAR with a lag.

[11] Real activity series correspond to categories 1–3 in the classification in Table 1 of Stock & Watson (2016), while price series correspond to category 6.

[12] The external IVs used in empirical practice tend to have low to moderate autocorrelation (Ramey, 2016), consistent with our assumptions on ρ_z.

DGP SUMMARY STATISTICS

Percentile	min	10	25	50	75	90	max
Data and shocks							
trace(long-run var)/trace(var)	0.42	0.93	0.98	1.14	2.29	4.78	18.09
Largest VAR eigenvalue	0.82	0.84	0.84	0.84	0.84	0.86	0.91
Fraction of VAR coef's $\ell \geq 5$	0.02	0.10	0.15	0.23	0.34	0.44	0.84
Degree of shock invertibility	0.14	0.16	0.19	0.28	0.41	0.47	0.65
IV first stage F-statistic	9.28	9.40	9.48	15.31	23.38	23.91	24.66
Impulse responses up to $h = 20$							
No. of interior local extrema	1	2	2	2	3	4	6
Horizon of max abs. value	0	0	0	0	1	2	8
Average/(max abs. value)	-0.42	-0.16	-0.08	-0.02	0.06	0.11	0.43
R^2 in regression on quadratic	0.01	0.09	0.20	0.46	0.69	0.83	0.97

Table 2.1: Quantiles of various population parameters across the 6,000 DGPs for observed shock and IV identification. "long-run var": long-run variance. "Fraction of VAR coef's $\ell \geq 5$": $\sum_{\ell=5}^{50} \|A_\ell^w\| / \sum_{\ell=1}^{50} \|A_\ell^w\|$, where $\|\cdot\|$ is the Frobenius norm. IV first stage F-statistic: $T \times R^2/(1-R^2)$, where $T = 200$ and R^2 is the population R-squared in a projection of i_t on $(z_t - E(z_t \mid z_{t-1}))$. "Average/(max abs. value)": $(\frac{1}{21}\sum_{h=0}^{20} \theta_h)/\max_h\{|\theta_h|\}$. "$R^2$ in regression on quadratic": R-squared from a regression of the impulse response function $\{\theta_h\}_{h=0}^{20}$ on a quadratic polynomial in h.

2.3.4 Summary statistics

Consistent with the experience of applied researchers, our DGPs exhibit substantial heterogeneity along several dimensions. Table 2.1 displays the distribution of various population parameters across our 6,000 DGPs. The table focuses on impulse responses with respect to directly observed monetary policy and government spending shocks, though results for recursively defined shocks are similar, as shown in Supplemental Appendix E.4.

First of all, the DGPs feature varying degrees of persistence. Our primary measure of the persistence of the DGP is given by $\text{trace}(LRV(\bar{w}_t))/\text{trace}(\text{Var}(\bar{w}_t))$, where "LRV" denotes the long-run variance matrix.[13] This measure varies widely across the DGPs, with more than 25% of the DGPs being anti-persistent on average (i.e., the ratio is less than 1), and the 90th percentile nearly equal to 5. The largest eigenvalue of the VAR companion matrix, another measure of persistence, has a median of 0.84 and does not vary as much across DGPs.[14] We consider an alternative set of more persistent DGPs in a robustness check in Section 2.5.5.

Second, the DGPs are heterogeneous in terms of how well they can be approximated by a relatively low-order VAR. Table 2.1 reports the ratio $\sum_{\ell=5}^{50} \|A_\ell^w\| / \sum_{\ell=1}^{50} \|A_\ell^w\|$, which measures the relative magnitude of the coefficient matrices $\{A_\ell^w\}_\ell$ in the VAR(∞) representation for $\{\bar{w}_t\}$ at or after lag 5 (with $\|\cdot\|$ denoting the Frobenius matrix norm). The 10th and 90th percentiles equal 0.10 and 0.44, respectively. Hence, the analysis in Section 2.2 suggests that the bias of low-order VAR procedures will vary substantially across the

[13] This measure equals $(1+\rho)/(1-\rho)$ for an AR(1) process with coefficient ρ.

[14] This is because the (transformed) interest rate and government spending series are quite persistent, and all DGPs include one of these two series.

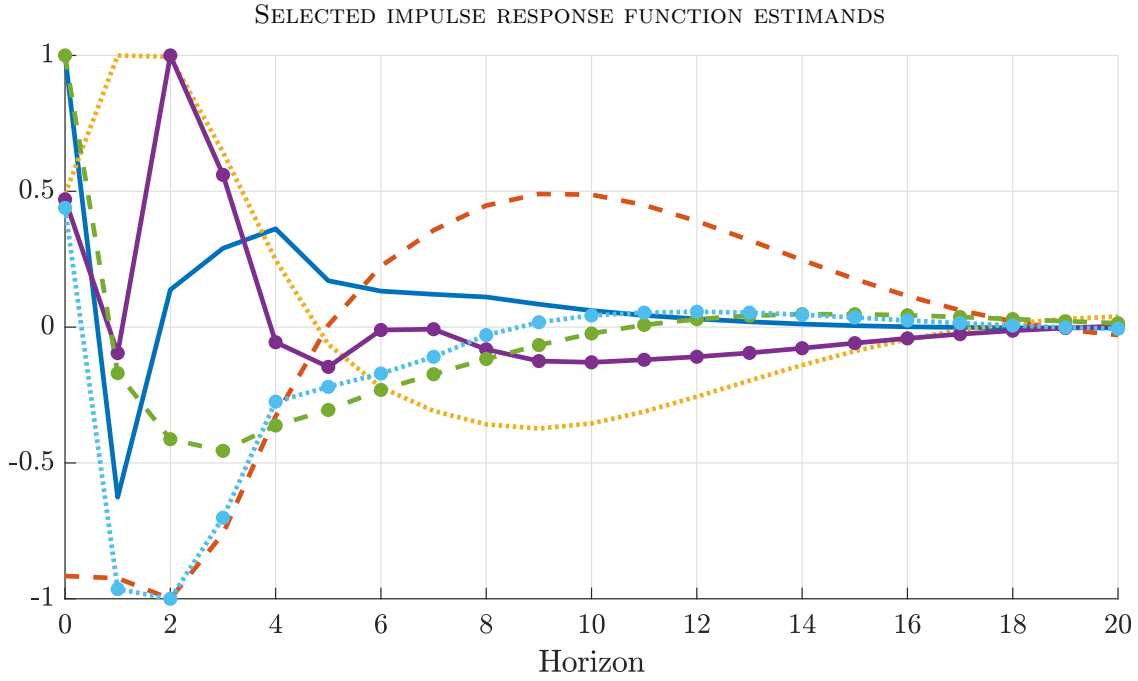


Figure 2.3: Selected impulse responses of macro observables to monetary and fiscal policy shocks. Here the impulse response functions are normalized to have a maximum value 1 or -1.

various DGPs that we consider in our simulations.

Third, for the IV specifications, the DGPs differ in terms of the degree of invertibility of the shock and the strength of the IV. The degree of invertibility is defined as the R-squared value in a population projection of the shock of interest on current and lagged macro observables $\{\bar{w}_{t-\ell}\}_{\ell=0}^{\infty}$. The bias of some SVAR-based external instrument procedures depends on how far below 1 this measure is, as discussed further in Section 2.4. The table shows that 90% of the DGPs have degrees of invertibility below 47%, i.e., substantial non-invertibility. This is not surprising: the DFM (2.5)–(2.7) features a realistic amount of idiosyncratic noise v_t, making it challenging to accurately back out the aggregate shock of interest $\varepsilon_{1,t}$ from a small number of observed time series $\bar{w}_t$. The strength of the IV is somewhat weak to moderate, as the population first stage F-statistic (from a regression of the normalization variable i_t on the IV z_t) varies between approximately 10 and 25, given sample size $T = 200$.

Finally, the true values of our impulse response estimands exhibit a wide variety of shapes. Table 2.1 shows that, though most impulse response functions peak at horizons $h = 0$ or $h = 1$, they are typically not simple monotonically decaying or even hump-shaped functions: the median number of interior local extrema of the impulse response functions is 2 (a monotonic function would have 0; a hump-shaped function would have 1). Many impulse response functions change sign at some horizon, as evidenced by the average response (across horizons) typically being much smaller than the maximal response. Finally, the smoothness of the impulse response functions varies substantially: the R-squared value in a regression of the impulse responses $\{\theta_h\}_{h=0}^{20}$ on a quadratic polynomial $b_0 + b_1 \times h + b_2 \times h^2$ has 10th and 90th percentiles given by 0.09 and 0.83, respectively. Figure 2.3 displays the true values of six impulse response functions that provide a representative picture of the heterogeneity. Note, however, that all of them tend to zero when the horizon

gets large, as required by the stationarity of the DGPs.

2.4 Estimation methods

We now give a brief overview of the different VAR- and LP-based estimation methods that we consider in the simulation study. Though all these methods aim at estimating the same population impulse responses defined in Section 2.3.2, the methods differ in terms of their bias/variance properties, and in terms of their robustness to non-invertibility. Implementation details are relegated to Section 2.B.[15] All estimators include an intercept.

LOCAL PROJECTION APPROACHES. The basic idea behind local projections, as proposed by Jordà (2005), is to estimate the impulse responses separately at each horizon by a direct regression of the future outcome on current covariates. We consider two such approaches:

1. **Least-squares LP.** OLS regression of the response variable y_{t+h} on the innovation variable x_t, controlling for p lags of all data series w_t. The innovation variable equals $x_t = \varepsilon_{1,t}$ for "observed shock" identification and $x_t = z_t$ for IV identification. In the case of recursive identification, we additionally control for the contemporaneous values of the variables that are ordered before x_t in the system (Plagborg-Møller & Wolf, 2021). Since least-squares LP does not mechanically impose any functional form on the relationship between impulse responses at different horizons h, the bias tends to be small. However, the estimated impulse response functions tend to look jagged in finite samples and tend to be estimated with high variance at longer horizons.

2. **Penalized LP** (abbreviated "Pen LP"). To lower the variance of least-squares LP at the expense of potentially increasing the bias, Barnichon & Brownlees (2019) propose a penalized regression modification of LP. The estimator minimizes the sum of squared forecast residuals (across both horizons and time) plus a penalty term that encourages the estimation of smooth impulse responses. This is a type of shrinkage estimation: the unrestricted least-squares estimate is pushed in the direction of a smooth quadratic function of the horizon. The degree of shrinkage is chosen by cross-validation.

In the case of IV identification, we apply the LP-IV estimation approach of Stock & Watson (2018), which is robust to non-invertibility.

VAR APPROACHES. Like local projections, a VAR with lag length p flexibly estimates the impulse responses out to horizon p; however, the VAR extrapolates the responses at longer horizons $h > p$ using only the sample autocovariances out to lag p. As suggested by the analysis in Section 2.2, this tends to generate impulse response estimates with lower variance but higher bias than LP estimates at intermediate horizons. In our stationary class of DGPs, VAR impulse response estimates eventually tend to zero when the horizon gets large (with high probability), unlike LP estimates; this feature will be important for the bias/variance trade-off at long horizons. We consider four VAR-based approaches:

[15] To visualize the various estimation methods, Supplemental Appendix D plots the estimated impulse response functions in a few data sets simulated from a single DGP.

1. **Least-squares VAR.** Standard VAR impulse response estimates based on equation-by-equation OLS estimates of the reduced-form coefficients.

2. **Bias-corrected VAR** (abbreviated "BC VAR"). As above, but using the formula in Pope (1990) to analytically correct the first-order bias of the reduced-form coefficients caused by persistent data.

3. **Bayesian VAR** (abbreviated "BVAR"). As above, but where the reduced-form coefficients are posterior mean estimates under a Minnesota-style prior.[16] Due to the stationarity of the DGP, we shrink towards independent white noise processes (rather than independent unit root processes, which is conventional for highly persistent data). The prior variance hyper-parameters follow the recommendations in Canova (2007).

4. **VAR model averaging** (abbreviated "VAR Avg"). Hansen (2016) develops a data-driven method for averaging across the impulse response estimates produced by several different VAR specifications. We construct a weighted average of 40 different specifications, each of which is estimated by OLS: univariate AR(1) to AR(20) models, and multivariate VAR(1) to VAR(20) models. The weights are chosen to minimize an empirical estimate of the final impulse response estimator's MSE.

   The VAR model averaging estimator effectively includes LP among the list of candidate estimators (as in the related approach of Miranda-Agrippino & Ricco, 2021a). This is because the candidate VAR(20) model gives results similar to LP with several lagged controls, at all horizons considered in our study (Plagborg-Møller & Wolf, 2021).

Observed shock identification is carried out by simply ordering the shock first in the recursive VAR. We consider two different approaches to IV estimation:

i) **Internal instruments.** Proceed as if the IV were equal to the true shock of interest, i.e., order the IV first in the VAR and compute responses to the first orthogonalized innovation (Ramey, 2011). Plagborg-Møller & Wolf (2021) prove that this approach consistently estimates the *normalized* structural impulse responses (2.10) even if the IV is contaminated with measurement error as in (2.9), and even if the shock is non-invertible.

ii) **SVAR-IV** (also known as proxy-SVAR). Exclude the IV from the reduced-form VAR, and estimate the structural shock by projecting the IV on the reduced-form VAR innovations (Stock, 2008; Stock & Watson, 2012; Mertens & Ravn, 2013; Gertler & Karadi, 2015). This estimator is consistent if the shock of interest is invertible, but not otherwise (Forni et al., 2019; Miranda-Agrippino & Ricco, 2021b; Plagborg-Møller & Wolf, 2022). We shall see that the SVAR-IV estimator tends to exhibit lower dispersion than the "internal instruments" estimator due to the smaller dimension of the VAR system, potentially justifying the use of the former despite its lack of robustness to non-invertibility.

We implement the "internal instruments" approach using all four types of VAR estimation techniques described earlier. For brevity, we only consider the least-squares version of the SVAR-IV estimator.

[16] We report posterior means of the impulse responses based on 100 draws.

LAG LENGTH SELECTION. As a baseline, the LP and VAR estimators use $p = 4$ lags for estimation (except VAR model averaging, which uses many different lag lengths). In our DGPs, the Akaike Information Criterion almost always selects very short lag lengths $\hat{p}_{AIC}$, as discussed further in Section 2.5.6 below. Thus, for all intents and purposes our results may be interpreted as having been generated by the lag length selection rule $p = \max\{\hat{p}_{AIC}, 4\}$. Our reading of applied practice is that researchers typically include at least 4 lags in quarterly data. Results for $p = 8$ are discussed in Section 2.5.5.

2.5 Results

This section presents our simulation results. We summarize the results through four takeaways, one in each subsection. The first three takeaways focus on observed shock identification. The fourth takeaway is concerned with IV identification. We show in Section 2.5.5 that these conclusions are qualitatively robust to several specification choices (including recursive identification). Finally, in Section 2.5.6, we justify our focus on the *average* performance of estimators across DGPs, by arguing that there is limited scope for selecting among estimators in a data-dependent way.

Throughout this section we present results for our 6,000 monetary and fiscal policy shock DGPs considered jointly rather than separately. For each DGP, we simulate time series of length $T = 200$ quarters and approximate the population bias and variance of the estimators by averaging across 5,000 Monte Carlo simulations. Our baseline results take about one week to produce in Matlab on 16 servers with 25 parallel cores each.

2.5.1 There is a clear bias-variance trade-off between LP and VAR

Our first takeaway is that researchers invariably face a bias-variance trade-off: because most of our DGPs are not well approximated by finite-order VAR models, least-squares LPs tend to have lower bias, while least-squares VAR estimators tend to have lower variance. This agrees with the simple analytical example in Section 2.2. We focus on the baseline least-squares LP and VAR estimators in this subsection, leaving other procedures for later.

Figures 2.4 and 2.5 depict the bias-variance trade-off at various horizons. These figures show the median (across our 6,000 DGPs) of the absolute bias or standard deviation, respectively, as a function of the horizon. The different lines correspond to different estimators, with least-squares LP and VAR being the thick lines. Before taking the median, we cancel out the units of the response variables by dividing the bias and standard deviation by $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$, i.e., the root mean squared value of the *true* impulse response function out to horizon 20. Note that the scale of the vertical axis differs between the bias and standard deviation plots.

The figures show that least-squares LP and VAR estimators have similar bias and variance at horizons $h \leq p = 4$, but not at longer horizons $h > p$. The median LP bias is close to zero at all horizons, yet the variance is high and does not decrease with the horizon.[17] In contrast, the median VAR variance decays quickly toward zero as a function of the horizon, at the cost of elevated bias at intermediate horizons (i.e., horizons h that are moderately larger than the estimation lag length p). These observations are consistent

[17] Kilian & Kim (2011) find in simulations that LP does not have lower bias than VAR estimators, but they consider a different variant of LP that uses an auxiliary VAR to identify the structural shocks.

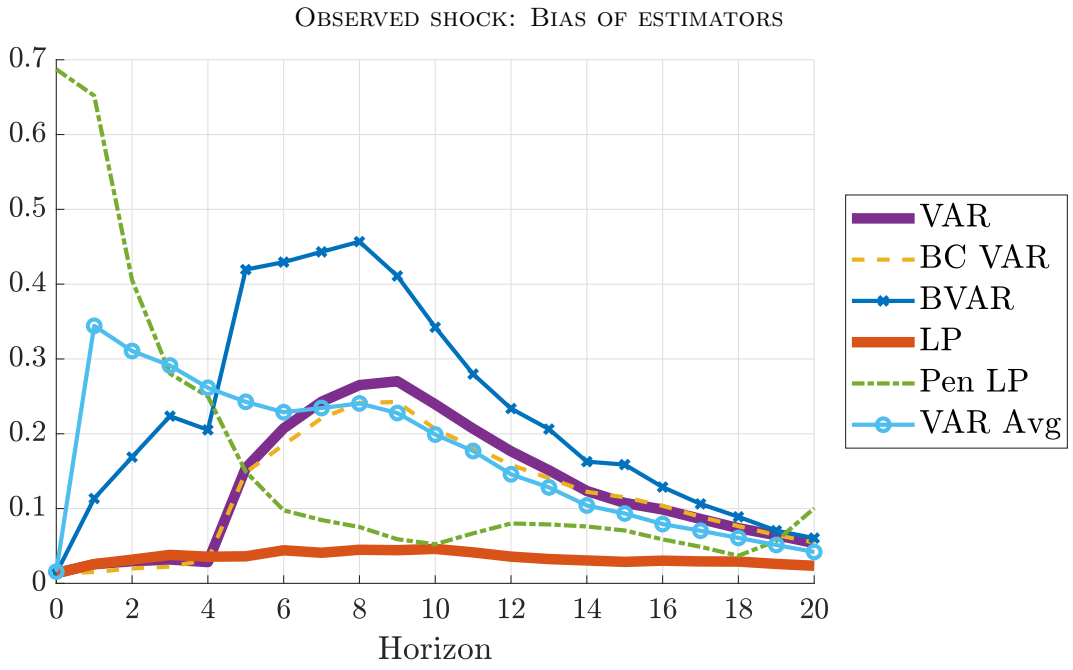


Figure 2.4: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

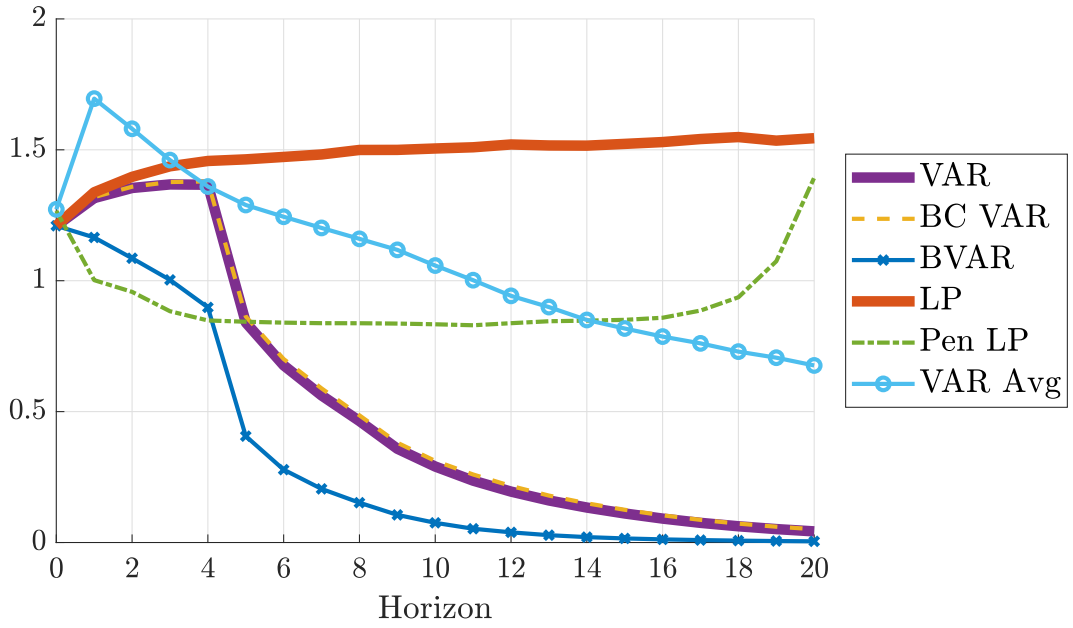


Figure 2.5: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

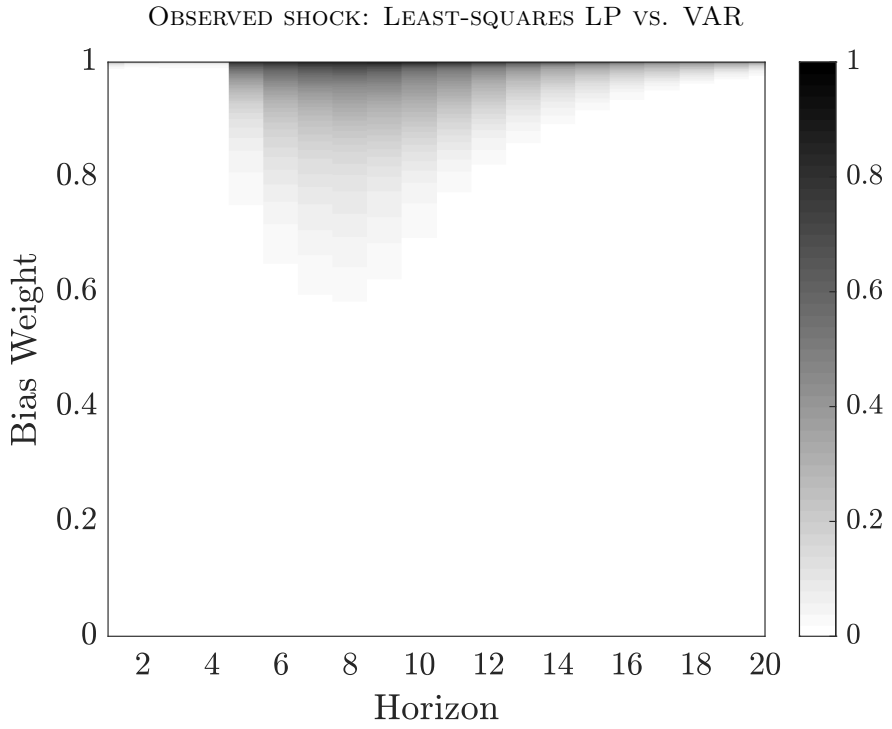


Figure 2.6: Fraction of DGPs for which the least-squares LP estimator has a lower loss than the VAR estimator. Darker areas correspond to regions where LP is preferred more often. Horizontal axis: impulse response horizon h. Vertical axis: weight ω on squared bias in the loss function (2.2). The loss function is normalized by the scale of the impulse response function, as in Figures 2.4 and 2.5. The impact horizon $h = 0$ is omitted due to numerical equivalence between the estimators.

with the theoretical results in Section 2.2, Schorfheide (2005), and Plagborg-Møller & Wolf (2021).

Figure 2.6 shows that least-squares VAR is preferred over least-squares LP for almost all horizons and almost all researcher loss functions; the only exceptions are when the weight ω on squared bias in the loss function (2.2) is near 1 and the horizon h is intermediate. The figure uses grey shadings to indicate the *fraction* of DGPs for which, at a given h and ω, the loss for least-squares LP is smaller than that for least-squares VAR. For $h \leq p = 4$, the VAR is preferred almost uniformly because its variance is slightly smaller; however, the difference is negligible. In the region with a non-trivial bias-variance trade-off—i.e., $h > 4$—a stark picture emerges: for the researcher to prefer the LP method, her relative weight on squared bias will usually have to be at least four times larger than that on variance. This is because, as already revealed by Figures 2.4 and 2.5, the relatively low bias of the LP method comes at a very steep cost in terms of increased sampling variance. Viewed through the lens of the analytical illustration in Figure 2.2, the findings coming out of our large number of empirically disciplined DGPs are consistent with a calibration of our simple model in Section 2.2 with (i) moderate persistence (ρ) and (ii) non-zero but limited VAR mis-specification (α).

2.5.2 Shrinkage dramatically lowers variance, at some cost of bias

Our second main takeaway is that shrinkage methods often allow a substantial reduction in variance, at a moderate cost in terms of bias. In this subsection we focus on two shrinkage methods in particular: penalized

LP, which shrinks the conventional least-squares LP estimates towards smooth impulse response functions, and the Bayesian VAR (BVAR) estimator, which shrinks the least-squares VAR coefficients towards independent white noise.

The dash-dotted and x-marked lines in Figure 2.4 show that penalized LPs and BVARs have uniformly higher bias than least-squares LPs and VARs, respectively. This is an inevitable cost of shrinkage, since the stylized prior information used for the shrinkage is never completely accurate. For penalized LP, the increase in bias is particularly stark at short horizons, since this estimator uses the longer-horizon impulse response estimates to inform the estimation of the short-run impulse responses (through the prior belief in smoothness).

Figure 2.5, however, shows that shrinkage can lead to large reductions in variance: the median standard deviation of penalized LP lies everywhere below that of standard LP (and is in fact the lowest of all methods for $h \leq p$), and the standard deviation of the BVAR estimator is always strictly below that of least-squares VAR.

Figures 2.7 and 2.8 depict the head-to-head loss-based comparison of penalized LP vs. least-squares LP, and of BVAR vs. least-squares VAR, respectively. Given the relatively high bias of penalized LP at short horizons, the trade-off for $h \leq p$ is non-trivial: the researcher's optimal choice of least-squares LP vs. penalized LP is sensitive to the bias weight ω. At intermediate horizons, however, the case for penalized LP becomes overwhelming: substantial decreases in variance are achieved at the cost of moderate increases in bias, so a preference for conventional LP methods requires an overwhelming concern for bias.[18] In the comparison of BVARs and least-squares VARs, the bias-variance trade-off is similarly non-trivial: shrinkage offers a very competitive reduction in variance in return for its increased bias, and so justification of least-squares VAR requires a relatively high bias weight ω. Note that if we focus on the longest horizons in Figures 2.4 and 2.5, the difference between least-squares VAR and BVAR is minimal, unlike when comparing the two LP procedures. This is because any kind of VAR-estimated impulse response function tends to zero as $h \to \infty$.

2.5.3 No method dominates, but shrinkage is generally welcome

Our third takeaway is that, for a fixed loss function, no single method is best at *all* horizons. Nevertheless, regardless of the horizon, one of the LP or VAR shrinkage procedures is preferred by most loss functions, unless bias receives a high weight.

Figure 2.9 shows the optimal estimation method as a function of the horizon h and the bias weight ω. We report the estimation method that minimizes the *average* loss (2.2) across DGPs, after normalizing the loss to cancel out units as in Figures 2.4 and 2.5.

The attractive performance of the shrinkage estimators is evident visually, as most of the figure is either cross-hatched green (penalized LP) or solid-dotted blue (BVAR). If interest centers solely on short impulse response horizons, penalized LP is almost always the best choice, except when ω is close to 1 (i.e., an almost exclusive concern for bias). At horizons longer than the lag length $p = 4$, some kind of VAR method is generally preferable. In particular, BVAR shrinkage is optimal if the weight on bias is not too large, while

[18] Penalized LP performs similarly to least-squares LP at horizon $h = 20$. The reason is mechanical: the Barnichon & Brownlees (2019) smoothness penalty affects estimates near the "boundary" (i.e., the largest penalized horizon) less than at the other horizons.

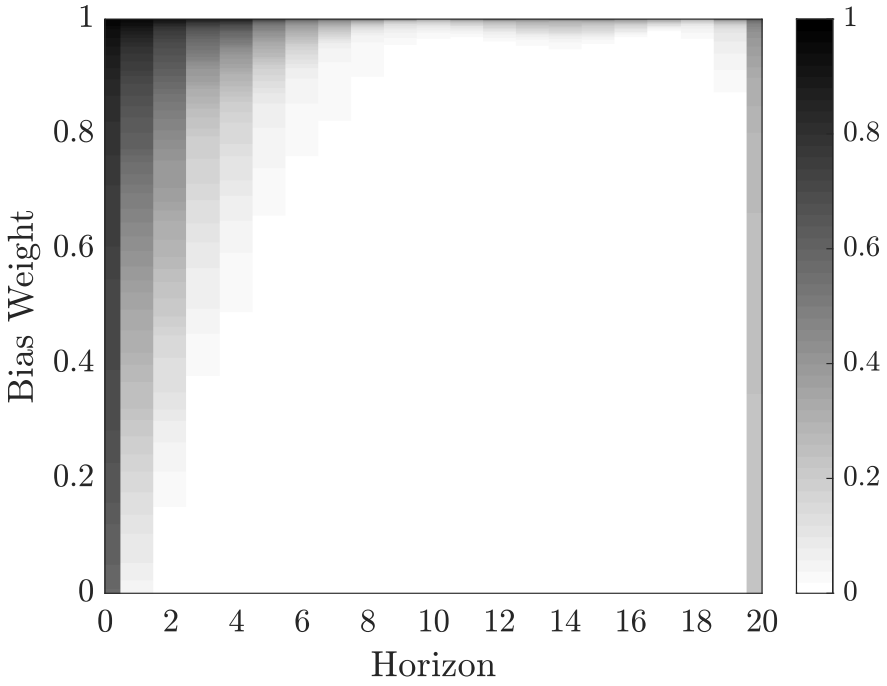


Figure 2.7: Fraction of DGPs for which the least-squares LP estimator has a lower loss than penalized LP. Darker areas correspond to regions where least-squares LP is preferred more often. See caption for Figure 2.6.

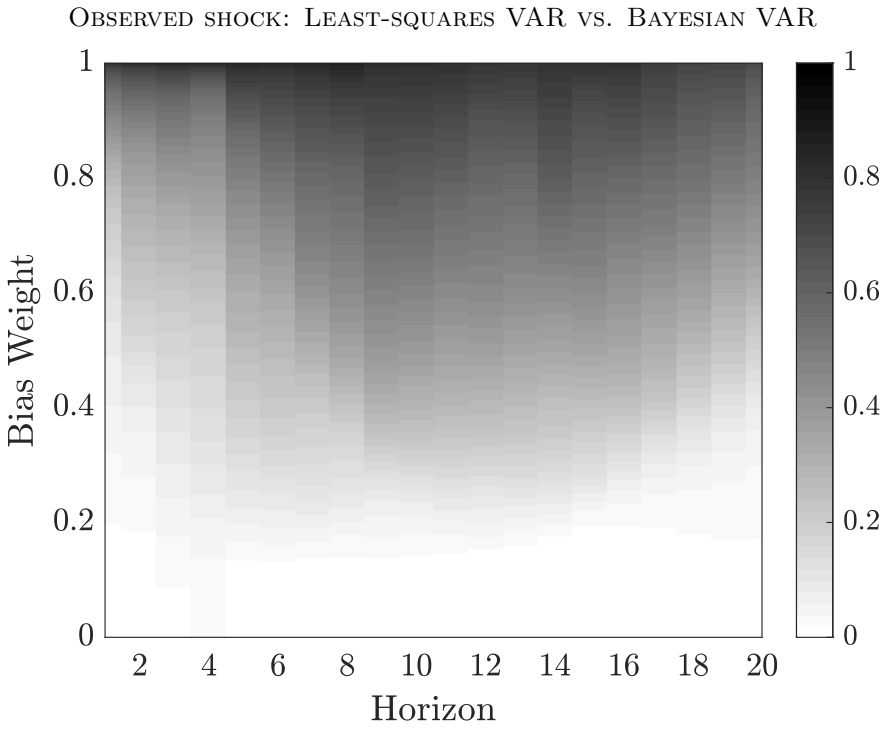


Figure 2.8: Fraction of DGPs for which the least-squares VAR estimator has a lower loss than the BVAR estimator. Darker areas correspond to regions where least-squares VAR is preferred more often. See caption for Figure 2.6. The impact horizon $h = 0$ is omitted due to numerical equivalence between the estimators.

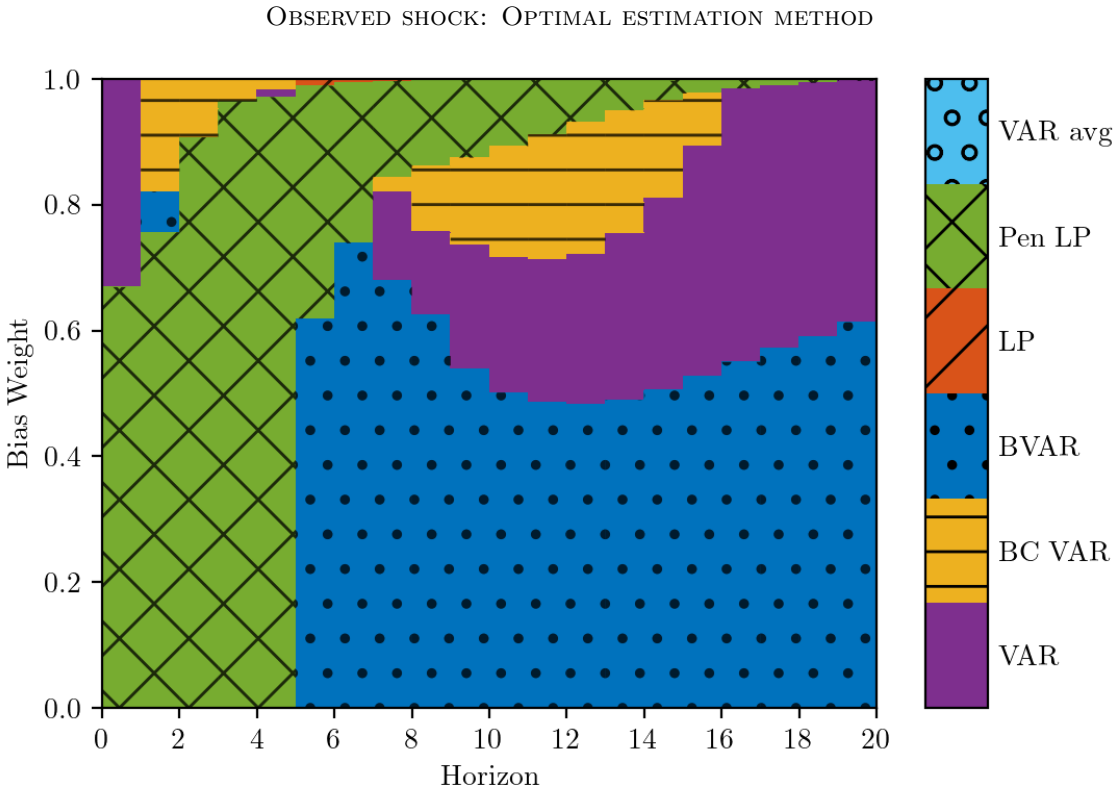


Figure 2.9: Method that minimizes the average (across DGPs) loss function (2.2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 2.4 and 2.5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

least-squares VAR or bias-corrected VAR are preferred if the weight on bias is high.

Though bias-corrected VAR (yellow with horizontal lines) looks attractive in Figure 2.9 when $\omega \geq 0.8$, Figure 2.4 shows that in practice the reduction in bias relative to the conventional least-squares VAR estimator is small. This finding is due to the modest persistence of our DGPs, as discussed in Section 2.3.4. Figure 2.9 also shows that least-squares LP (orange with diagonal lines) as well as the VAR model averaging estimator (light blue with hollow circles) are almost never optimal.[19] Least-squares LP is essentially dominated because its low bias is not much lower than that of penalized LP at intermediate horizons, and so the high variance of least-squares LP is difficult to justify. The reason why VAR model averaging does poorly is due to its high median standard deviation relative to other VAR-based estimators (see Figure 2.5). Closer inspection reveals that the high standard deviation is a consequence of the estimator having a very fat-tailed sampling distribution, with a non-negligible probability of erratic estimates.[20]

[19] Least-squares LP is in fact strictly preferred in a very thin region with intermediate p and $\omega \approx 1$.

[20] We use Hansen's (2016) code off the shelf. It would be interesting to investigate whether the procedure could be modified to avoid erratic estimates.

2.5.4 SVAR-IV is heavily biased, but has relatively low dispersion

Our last takeaway is concerned with IV/proxy estimation procedures. Among the robust "internal instruments" procedures, the bias-variance trade-off is very similar to that discussed above for the case of an observed shock. The "external instruments" SVAR-IV procedure, however, contributes very starkly to the trade-off: it can be severely biased due to its lack of robustness to non-invertibility, but at the same time it has substantially lower dispersion than the "internal instruments" procedures.

Figures 2.10 and 2.11 show the median bias and interquartile range of the various IV estimators. We here report median bias instead of (mean) bias and the interquartile range instead of standard deviation, because the sampling distributions of the IV impulse response estimates is fat-tailed, as is often the case with moderately strong IVs.[21] If we ignore the dotted burgundy line for SVAR-IV, these figures are qualitatively similar to those presented in Section 2.5.1. However, SVAR-IV stands out by exhibiting especially high median bias at short horizons, but low interquartile range at short and intermediate horizons. This is consistent with the existing theoretical work referenced in Section 2.4: unlike the "internal instruments" procedures, SVAR-IV is asymptotically biased when the shock is not invertible, and we saw in Section 2.3.4 that the degree of invertibility is generally low in our DGPs. On the other hand, the SVAR-IV procedure has fewer parameters to estimate, as it excludes the IV z_t from the reduced-form VAR regression, causing a reduction in dispersion relative to the other procedures. Though we view the high median bias of SVAR-IV across our DGPs as a worrying finding, its low dispersion is intriguing and may in some cases in fact trump the bias concerns.

2.5.5 Robustness

This section argues that our main conclusions in Sections 2.5.1 to 2.5.4 are robust to several alterations of our baseline simulation specification. Since our baseline DGPs are not highly persistent (see Section 2.3.4), we here pay particular attention to robustness exercises that consider DGPs with greater persistence. Various other robustness checks are listed subsequently, with details relegated to Supplemental Appendix E.

Persistence. We consider an alternative set of DGPs with higher persistence than our baseline specification. This is achieved by adjusting the VAR(2) process for the factors f_t so that the largest absolute eigenvalue of the VAR polynomial equals 0.95, implying that some impulse responses in each DGP decay with a half-life of approximately 3.4 years. We correspondingly scale the factor innovations to keep the trace of the factor variance-covariance matrix constant. See Section 2.A.3 for details on the adjustment.

Higher persistence does not change our qualitative conclusions, except that the performance of the Bayesian VAR procedure becomes more sensitive to the prior at long horizons. Figure 2.12 shows the optimal method choice for this more persistent set of DGPs; other results are presented in Supplemental Appendix E.2. At short horizons, the results are similar to the baseline results in Figure 2.9, with penalized LP an attractive option unless the concern for bias is overwhelming. At long horizons, the least-squares or bias-corrected VAR procedures do relatively better than the Bayesian VAR procedure compared to our baseline. This is because the BVAR procedure we consider shrinks towards white noise, which incurs a

[21] For completeness, (mean) bias and standard deviation are reported in Supplemental Appendix E.1.

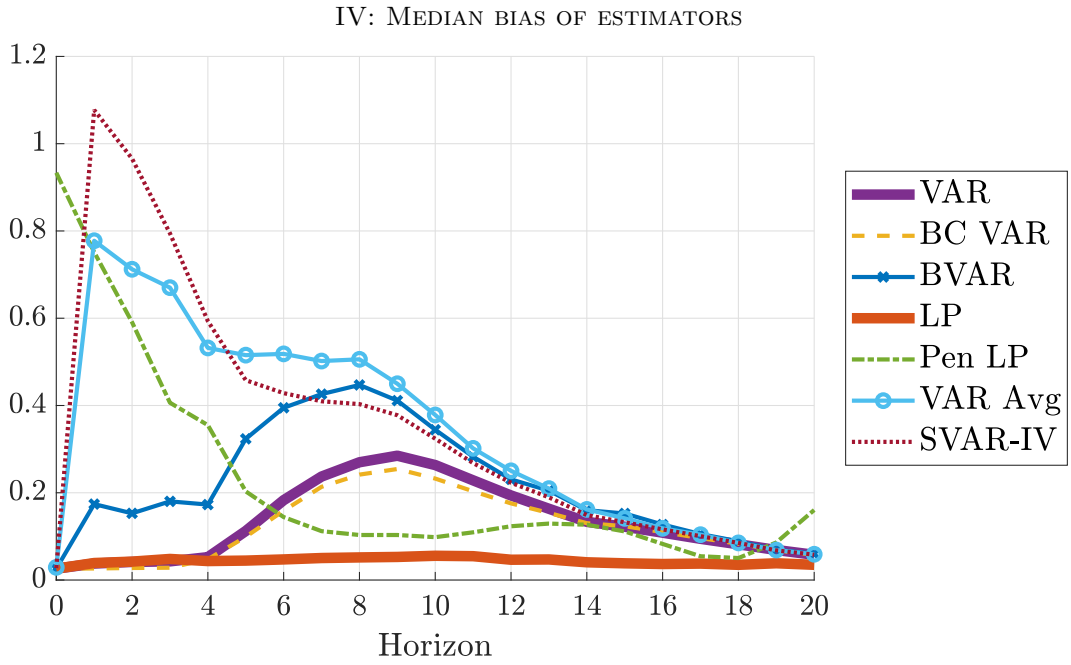


Figure 2.10: Median (across DGPs) of absolute median bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

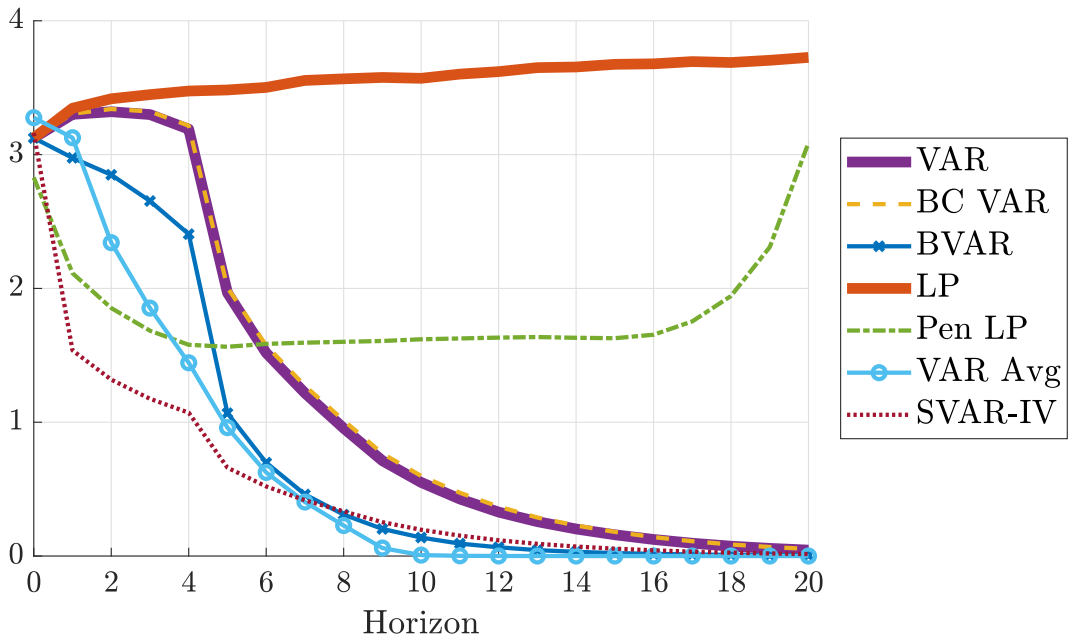


Figure 2.11: Median (across DGPs) of interquartile range of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

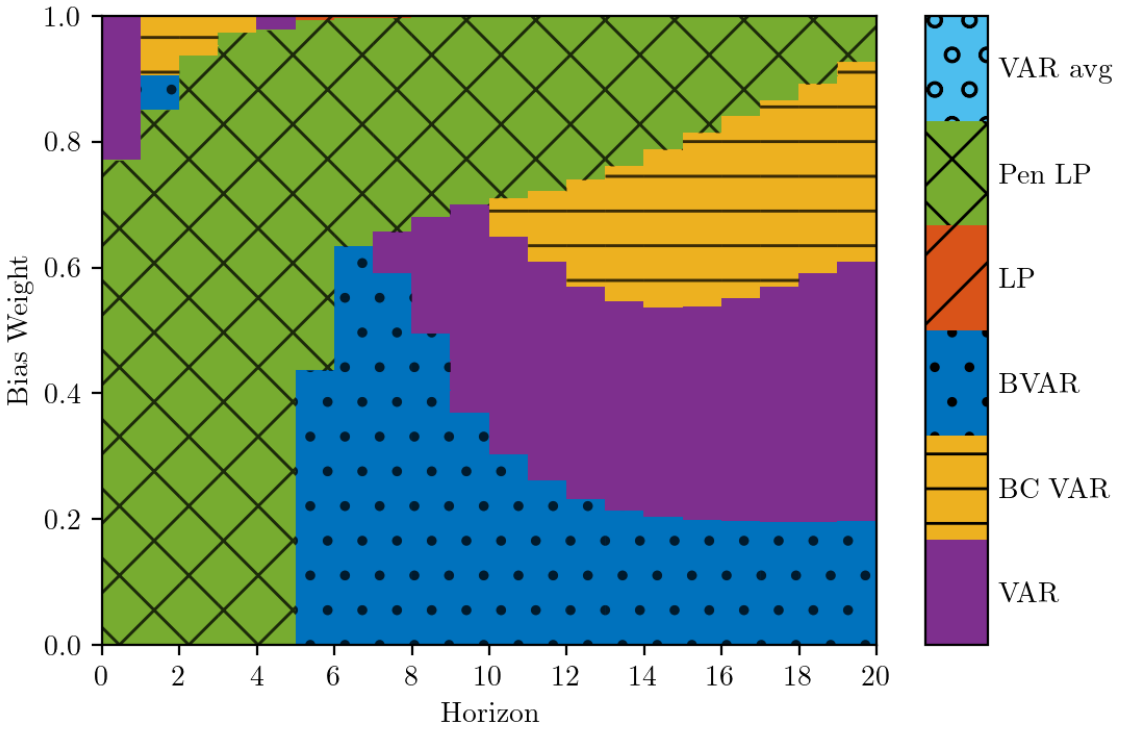


Figure 2.12: Method that minimizes the average (across DGPs) loss function (2.2) for the alternative persistent DFM. Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 2.4 and 2.5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

substantial bias for some persistent DGPs. The natural alternative of centering the prior at independent random walks produces only slightly lower bias. We thus conclude that, when the data is very persistent and interest centers on long horizons, prior elicitation for Bayesian VARs should receive careful attention and should not be left to convention. Another finding is that the bias correction of the VAR estimator has more bite in the persistent set of DGPs, though the performance of this estimator remains qualitatively comparable to the un-corrected least-squares VAR estimator.

Rather than directly changing the persistence of the DGPs, we also consider an alternative exercise that *cumulates* impulse responses across horizons. Starting with our baseline results, we compute cumulated impulse responses for those time series in the Stock & Watson (2016) data set that have been differenced prior to estimation. Thus, the resulting impulse responses can be interpreted as the responses of the levels of these variables. With this change in estimand, both bias and variance of the estimates now tend to increase with the horizon, as shown in Supplemental Appendix E.3. Yet our takeaways regarding the *relative* magnitudes of bias and variance are not materially affected. In particular, unless the concern for bias is substantial, the bias-variance trade-off continues to be best addressed by employing shrinkage techniques such as penalized LPs or BVARs.

OTHER ROBUSTNESS CHECKS. The following modifications to our baseline simulation specification all leave our main conclusions qualitatively unchanged.

- **Recursive identification:** Results for recursively identified shocks are strikingly similar to those for observed shock identification. See Supplemental Appendix E.4.
- **Monetary vs. fiscal shocks:** If we consider the monetary shock DGPs separately from the fiscal shock DGPs, the bias-variance trade-off is qualitatively similar to when we consider the DGPs jointly. See Supplemental Appendix E.5.
- **Lag length:** If the estimation lag length p is set to 8 instead of 4, our conclusions go through as long as we reinterpret "short horizons" to mean $h \leq 8$ instead of $h \leq 4$. See Supplemental Appendix E.6.
- **Sample size:** Halving the sample size to $T = 100$ quarters tends to increase the estimator standard deviations more than the biases, so shrinkage techniques look even more desirable than in our baseline. See Supplemental Appendix E.7.
- **Variable categories:** We find little evidence that the biases or standard deviations of individual impulse response estimators depend systematically on which categories of time series are included in the DGP (e.g., how many financial variables are used). See Supplemental Appendix E.8.
- **Salient observables:** Our results remain essentially unchanged if we restrict attention to a subset of 18 oft-used, salient macroeconomic time series out of the 207 ones in the Stock & Watson (2016) data set. We consider the exhaustive list of *all* five-variable DGPs that can be formed from these 18 series, subject to the selection rules in Section 2.3.3. See Supplemental Appendix E.9.
- **Near-worst-case performance:** Whereas our baseline results pertain to the median performance of estimators across DGPs, some researchers may instead prefer to focus on ensuring acceptable performance for particularly challenging DGPs. To this end, Supplemental Appendix E.10 reports the 90th percentiles of the bias and standard deviation across DGPs. Interestingly, adopting this "near-worst-case" perspective does not alter much the *relative* magnitudes of bias and standard deviation across estimation procedures. Hence, none of the estimation procedures seem to have a particular advantage in ensuring robustness to challenging environments, over and above their performance on typical DGPs.

2.5.6 Discussion: can we select the estimator based on the data?

Whereas the preceding results have focused on the performance of estimators *on average* across our many DGPs, it is natural to ask whether the choice of estimator can be guided by the data at hand. We now show that this appears to be difficult, as conventional model selection or evaluation criteria are unable to detect even substantial mis-specification of the VAR(4) model in the vast majority of our DGPs. These findings are consistent with the previously documented poor performance of the VAR model averaging estimator. For simplicity, we focus again on the "observed shock" DGPs.

First, the Akaike Information Criterion (AIC) tends to select very short lag lengths $\hat{p}_{AIC}$ in our DGPs, as mentioned earlier. The 90th percentile of $\hat{p}_{AIC}$ (across simulations) does not exceed 2 in any of our 6,000

DGPs, and it equals 2 in only 11.0% of the DGPs. This frequently used model selection tool therefore essentially never indicates that the VAR(4) specification is mis-specified.

Second, the Lagrange Multiplier test of residual serial correlation has low power in most of our DGPs. We carry out this test by regressing the VAR sample residuals on their first lags, controlling for four lags of the observed variables, and employing the specific likelihood ratio test defined in Johansen (1995). Using a 10% significance level for the test, less than 0.1% of the DGPs exhibit a rejection probability above 25%. Hence, this conventional specification test of the VAR(4) model is severely under-powered, despite the fact that many of our DGPs are in fact not well approximated by a VAR(4) model in population (see Section 2.3.4).

It is of course possible that other model selection criteria or specification tests will work better. However, at a minimum, the performance of the VAR model averaging estimator discussed in Section 2.5.3 and the evidence presented in this subsection together suggest that it is not straightforward to develop effective data-dependent estimator selection rules for use on conventional macroeconomic time series data. We therefore leave the issue of data-driven estimator selection to future research.

2.6 Conclusion and directions for future research

We conducted a large-scale simulation study of the performance of LP and VAR structural impulse response estimators, as well as several variants of these methods. By generating thousands of DGPs from an empirically estimated encompassing dynamic factor model, we ensured that our results apply to a wide range of settings faced by applied researchers. We drew the following four conclusions.

1. As predicted by theory, there is a non-trivial bias-variance trade-off between least-squares LP and VAR estimators. Empirically relevant DGPs are unlikely to admit finite-order VAR representations, so mis-specification of VAR estimators is a valid concern, as discussed by Ramey (2016) and Nakamura & Steinsson (2018). Reducing this bias by going all the way to LP estimation, however, tends to come at a surprisingly steep cost in terms of increased sampling variance at intermediate and long horizons.

2. Shrinkage procedures—such as the penalized LP procedure of Barnichon & Brownlees (2019) and Bayesian VAR estimation—can dramatically lower the variance at the cost of often only moderately larger bias. Typically the researcher must care overwhelmingly about bias to justify the use of conventional least-squares estimators over shrinkage procedures. An important caveat is that, if the data is highly persistent and interest centers on long horizons, shrinkage via Bayesian VARs requires careful prior elicitation. The literature on shrinkage estimators of impulse responses is not yet saturated, and we believe our findings are encouraging for further work in this area.

3. No method dominates at all response horizons, though one of the shrinkage procedures is usually optimal or near-optimal. Penalized LP is especially attractive at short horizons, while BVAR estimation is generally attractive—or at least competitive—at intermediate and long horizons. One might hope to achieve the best of both worlds by selecting among estimators in a data-dependent way, but preliminary evidence suggests that the scope for this is limited. In particular, we find that conventional model selection or evaluation tools are unable to detect substantial VAR mis-specification in realistic sample sizes. Despite our negative results in this area, we view data-dependent estimator selection as an important topic for future research.

4. In the case of IV identification, the popular SVAR-IV (or proxy-SVAR) procedure can be badly biased at short horizons, but it has substantially lower dispersion at all horizons than "internal instruments" procedures such as LP-IV. The high (median) bias of SVAR-IV is due to its lack of robustness to non-invertibility, which is a pervasive and realistic feature of our DGPs. An interesting question is whether it is possible to develop alternative non-invertibility-robust estimation procedures that capture some of the variance improvement enjoyed by SVAR-IV.

Whereas our simulation DGPs are calibrated to stationarity-transformed macroeconomic time series, future research could fruitfully analyze the bias-variance trade-off in other data environments. First, due to space constraints, we have not considered applying the estimation procedures directly to time series data with unit roots. Second, we conjecture that the nature of the bias-variance trade-off may differ somewhat in panel data settings, to the extent that the availability of a large cross section reduces the sampling variance of the estimators for a given time dimension.

Appendix 2.A Details on DGP definitions

2.A.1 Shock definition

Our definition of the structural shock of interest, $\varepsilon_{1,t}$, ensures that it has the largest possible contemporaneous effect on nominal interest rates (for monetary shocks) and government spending (for fiscal shocks). Letting $\eta_t = H\varepsilon_t$ and ι^* denote the index of the policy instrument, the shock is thus defined through the solution of the following problem:

$$\max_{H} \quad \bar{\Lambda}_{\iota^*,\bullet} H e_1 \quad \text{s.t.} \quad HH' = \Sigma_\eta,$$

where $\Sigma_\eta \equiv \text{Var}(\eta_t)$, and e_1 selects the first column of H. The solution is given by $H_{\bullet,1} = \Sigma_\eta \bar{\Lambda}'_{\iota^*,\bullet}(\bar{\Lambda}_{\iota^*,\bullet}\Sigma_\eta \bar{\Lambda}'_{\iota^*,\bullet})^{-1/2}$.[22]

2.A.2 IV process calibration

We calibrate the parameters of the IV equation (2.9) using the shock series of Romer & Romer (2004) and Ben Zeev & Pappa (2017) for monetary and fiscal simulation DGPs, respectively. We regress each shock series $\hat{z}_t$ on leads and lags of the estimated factor innovations $\hat{\eta}_t$:

$$\hat{z}_t = \hat{c} + \sum_{l=-p}^{p} \hat{a}'_l \hat{\eta}_{t-l} + \text{residual}_t.$$

We choose the number p of leads and lags based on the BIC. If the underlying structural shock measured by $\hat{z}_t$ is recoverable with respect to the factors in the DFM (Plagborg-Møller & Wolf, 2022), then the R^2 in the above two-sided regression will pin down the signal-to-noise ratio of $\hat{z}_t$. We calibrate the signal-to-noise ratio of the simulation IV (2.9) to be consistent with this estimated signal-to-noise ratio. Specifically, $\sigma^2_\nu = \frac{1}{R^2} - 1$.

2.A.3 More persistent factor model

The more persistent encompassing factor model studied in Section 2.5.5 adjusts the baseline factor model as follows. We replace the polynomial $\Phi(L) = \Phi_1 + \Phi_2 L$ in the VAR(2) process (2.5) for the factors by the polynomial $\tilde{\Phi}(L) = \tilde{\Phi}_1 + \tilde{\Phi}_2 L$, where $\tilde{\Phi}_1 \equiv b \times \Phi_1$, $\tilde{\Phi}_2 \equiv b^2 \times \Phi_2$, $b \equiv 0.95/\lambda_{\max}$, and $\lambda_{\max}$ is the absolute value of the largest eigenvalue of the original VAR polynomial $I - L\Phi(L)$. It is straightforward to verify that this adjustment scales all eigenvalues up by a factor of b. In particular, the largest absolute eigenvalue of $I - L\tilde{\Phi}(L)$ equals 0.95. To hold constant the explanatory power of the factors in the DFM, we multiply the matrix H by a scalar factor such that the trace of $\text{Var}(f_t)$ is left unchanged.

[22] The remaining columns in H are chosen arbitrarily to satisfy the variance-covariance constraint; these columns only matter for the simulation results through the implications for reduced-form dynamics.

Appendix 2.B Details on estimation procedures

LEAST-SQUARES LP. The least-squares LP estimator of the impulse response at horizon h is based on the coefficient $\hat{\beta}_h$ in the h-step-ahead OLS regression

$$y_{t+h} = \hat{\mu}_h + \hat{\beta}_h x_t + \hat{\zeta}_h q_t + \sum_{\ell=1}^{p} \hat{\varphi}_{h,\ell} w_{t-\ell} + \text{residual}_h, \tag{2.B.1}$$

that is, we regress on the variable x_t, with controls given by the vector q_t as well as p lags of all of the data w_t. The estimands of Section 2.3.2 can now be estimated as follows:

1. **Observed shock.** We set x_t equal to the observed shock $\varepsilon_{1,t}$ and omit the contemporaneous controls q_t (we still control for lagged data).[23]

2. **IV.** We estimate a Two-Stage Least Squares (2SLS) version of (2.B.1), setting x_t equal to the normalization variable i_t, and instrumenting for this variable with the IV z_t. We omit q_t in this specification (but still include lagged controls). This is numerically the same as doing a LP of y_{t+h} on z_t (with lagged controls), and dividing this coefficient by the LP coefficient in a regression of i_t on z_t (with lagged controls), see Stock & Watson (2018) and Plagborg-Møller & Wolf (2021).

3. **Recursive identification.** x_t is the innovation variable, while q_t are the variables ordered before x_t in the identification scheme (Plagborg-Møller & Wolf, 2021).

PENALIZED LP. The Barnichon & Brownlees (2019) estimator lowers the variance of LP by exploiting a prior belief in smoothness of the impulse response function across horizons. Following their preferred implementation, we model the impulse response function using B-spline basis functions. The jaggedness penalty function penalizes deviations from a quadratic function of the horizon h. We penalize impulse responses up to horizon $\bar{h} = 20$. The penalty parameter is selected in a data-dependent way using 5-fold cross-validation. We do not penalize the coefficients on the control variables in the LP. When reporting relative impulse responses (2.10), we divide by the *least-squares* LP estimate of the impact response of the normalization variable i_t to the structural shock.

LEAST-SQUARES VAR. The least-squares VAR coefficient estimates are obtained through equation-by-equation OLS regressions. We perform a conventional Cholesky decomposition of the forecast error variance-covariance matrix and compute impulse response functions with respect to the orthogonalized shocks. The estimands of Section 2.3.2 can now be estimated as follows:

1. **Observed shock.** The shock $\varepsilon_{1,t}$ is ordered first in w_t, and we compute responses to the first innovation.

2. **IV.** We initially consider an "internal instruments" approach as in Ramey (2011). That is, we include the IV z_t in the data vector w_t, order the IV first, and compute responses with respect to the first

[23] The lags are not needed for consistency in this case, but they often improve efficiency.

innovation (Plagborg-Møller & Wolf, 2021). The relative impulse response (2.10) is obtained by dividing by the impact response of the normalization variable i_t.

3. **Recursive identification.** The ordering of variables in w_t equals the ordering of the desired population impulse response estimand (cf. Section 2.3.2), and we compute responses to the innovation of the policy instrument i_t.

In contrast to the above internal instruments approach, the SVAR-IV (or "proxy-SVAR") estimator of Stock (2008) is obtained by computing the reduced-form impulse responses $\hat{\Psi}_h$ $(h = 0, 1, \dots)$ corresponding to a VAR in $\bar{w}_t$ (i.e., excluding z_t), and then reporting relative impulse responses (2.10) corresponding to the absolute structural impulse responses $\hat{\Psi}_h \hat{\gamma}$, where $\hat{\gamma}$ is the sample covariance vector of the reduced-form VAR residuals $\hat{u}_t$ and the IV z_t.

Bayesian VAR. Our BVAR implementation assumes a Gaussian VAR(p) model with a Gaussian prior on the VAR coefficients (the residual variance-covariance matrix $\hat{\Sigma}$ is fixed at the unconstrained least-squares estimate). The prior is a version of the popular "Minnesota prior". Following Kilian & Lütkepohl (2017, chapter 5.2.3), we adapt this prior to a stationary setting by centering the prior of the VAR reduced-form coefficients at zero, which would imply that the time series are independent white noise processes. The prior variance hyper-parameters are given by the default choices in Canova (2007, chapter 10.2.2).[24]

VAR model averaging. Hansen (2016) proposes a data-dependent procedure for averaging across impulse responses estimates produced by a collection of different AR and VAR models with different lag lengths. Let $\hat{\delta}_h(r)$ denote the un-normalized, least-squares recursive impulse response estimate at some horizon h for model $r = 1, \dots, R$. We estimate $\hat{\delta}_h(r)$ from $R = 40$ candidate models: First, univariate AR models for y_t with lag lengths from $p = 1$ up to $p = 20$; and second, VAR models in w_t with lag lengths from $p = 1$ up to $p = 20$. As in Hansen (2016), the variance-covariance matrix of innovations Σ and thus the impact effect δ_0 are fixed across candidate models and treated as known without error.[25] The VAR model averaging estimator is given by $\sum_{r=1}^{R} \hat{\omega}_r \hat{\delta}_h(r)$, where the weights $\{\hat{\omega}_r\}_{r=1}^{R}$ are chosen to minimize the data-dependent approximated MSE estimate $\hat{M}(\omega_1, \dots, \omega_R) \approx E[T(\sum_{r=1}^{R} \omega_r \hat{\delta}_h(r) - \delta_h)^2]$, subject to the constraints that all weights are nonnegative and $\sum_{r=1}^{R} \omega_r = 1$. Details of the MSE estimate are given in Hansen (2016, Section 6).[26] We run this optimization for the weights separately at each impulse response horizon. Relative impulse responses (2.10) are computed by dividing the absolute impulse response by the least-squares VAR(4) impact impulse response estimate of i_t with respect to the identified shock.

[24] Specifically, for each lag ℓ, the prior variance is $0.04/\ell^2$ for lags of the same variable, $0.01/\ell^2$ for lags of other variables, and 4000 for the intercept.

[25] To match the impact effect estimate in our benchmark method of least-squares VAR, we use $\hat{\Sigma}$ from the $p = 4$ VAR estimate as the true value across all the candidate models.

[26] The object of interest, $\hat{\delta}_h(r)$, is a scalar, which allows us to omit the weighting matrix required in the MSE estimate in Hansen (2016).

References

Austin, B. A. (2020). *Essays on Labor Economics and Econometrics*. PhD thesis, Harvard University. Chapter 2: "The trade-off between LP-IV and SVAR-IV estimation".

Barnichon, R. & Brownlees, C. (2019). Impulse Response Estimation by Smooth Local Projections. *The Review of Economics and Statistics*, *101*(3), 522–530.

Ben Zeev, N. & Pappa, E. (2017). Chronicle of a war foretold: The macroeconomic effects of anticipated defence spending shocks. *The Economic Journal*, *127*(603), 1568–1597.

Blanchard, O. & Perotti, R. (2002). An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *Quarterly Journal of Economics*, *117*(4), 1329–1368.

Brugnolini, L. (2018). About Local Projection Impulse Response Function Reliability. CEIS Research Paper, Vol. 16, Issue 6, No. 440.

Bruns, M. & Lütkepohl, H. (2021). Comparison of Local Projection Estimators for Proxy Vector Autoregressions. DIW Berlin Discussion Paper 1949.

Canova, F. (2007). *Methods for Applied Macroeconomic Research*. Princeton University Press.

Choi, C.-Y. & Chudik, A. (2019). Estimating impulse response functions when the shock series is observed. *Economics Letters*, *180*, 71–75.

Christiano, L., Eichenbaum, M., & Evans, C. (1999). Monetary Policy Shocks: What Have We Learned and to What End? In J. B. Taylor & M. Woodford (Eds.), *Handbook of Macroeconomics*, volume 1A chapter 2, (pp. 65–148). Elsevier.

Forni, M., Gambetti, L., & Sala, L. (2019). Structural VARs and noninvertible macroeconomic models. *Journal of Applied Econometrics*, *34*(2), 221–246.

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

Hansen, B. E. (2016). Stein Combination Shrinkage for Vector Autoregressions. Manuscript, University of Wisconsin-Madison.

Inoue, A. & Kilian, L. (2020). The uniform validity of impulse response inference in autoregressions. *Journal of Econometrics*, *215*(2), 450–472.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

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

Kilian, L. (1998). Small-sample Confidence Intervals for Impulse Response Functions. *Review of Economics and Statistics*, *80*(2), 218–230.

Kilian, L. & Kim, Y. J. (2011). How Reliable Are Local Projection Estimators of Impulse Responses? *Review of Economics and Statistics*, *93*(4), 1460–1466.

Kilian, L. & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Lazarus, E., Lewis, D. J., Stock, J. H., & Watson, M. W. (2018). HAR Inference: Recommendations for Practice. *Journal of Business & Economic Statistics*, *36*(4), 541–559.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1–2), 499–526.

Meier, A. (2005). How Big is the Bias in Estimated Impulse Responses? A Horse Race between VAR and Local Projection Methods. Manuscript, European University Institute.

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

Miranda-Agrippino, S. & Ricco, G. (2021a). Bayesian Local Projections. Warwick Economics Research Papers 1348.

Miranda-Agrippino, S. & Ricco, G. (2021b). Identification with External Instruments in Structural VARs. Manuscript, University of Warwick.

Montiel Olea, J. L. & Plagborg-Møller, M. (2021). Local Projection Inference Is Simpler and More Robust Than You Think. *Econometrica*, *89*(4), 1789–1823.

Nakamura, E. & Steinsson, J. (2018). Identification in Macroeconomics. *Journal of Economic Perspectives*, *32*(3), 59–86.

Plagborg-Møller, M. & Wolf, C. K. (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, *89*(2), 955–980.

Plagborg-Møller, M. & Wolf, C. K. (2022). Instrumental Variable Identification of Dynamic Variance Decompositions. *Journal of Political Economy*, *130*(8), 2164–2202.

Pope, A. L. (1990). Biases of Estimators in Multivariate Non-Gaussian Autoregressions. *Journal of Time Series Analysis*, *11*(3), 249–258.

Ramey, V. A. (2011). Identifying Government Spending Shocks: It's All in the Timing. *Quarterly Journal of Economics*, *126*(1), 1–50.

Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation. In J. B. Taylor & H. Uhlig (Eds.), *Handbook of Macroeconomics*, volume 2 chapter 2, (pp. 71–162). Elsevier.

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

Schorfheide, F. (2005). VAR forecasting under misspecification. *Journal of Econometrics*, *128*(1), 99–136.

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

Stock, J. H. (2008). What's New in Econometrics: Time Series, Lecture 7. Lecture slides, NBER Summer Institute.

Stock, J. H. & Watson, M. W. (2012). Disentangling the Channels of the 2007–09 Recession. *Brookings Papers on Economic Activity*, *2012*(1), 81–135.

Stock, J. H. & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, volume 2 chapter 8, (pp. 415–525). Elsevier.

Stock, J. H. & Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal*, *128*(610), 917–948.

Chapter 3

International Evidence on Credit and Economic Activity

3.1 Introduction

There has been longstanding arguments about whether credit expansion is good or bad for economic growth after 2008.[1] Conventional wisdom usually regards a big credit-to-GDP ratio as a signal for output boost, when comparing income levels across countries. Classical textbooks also treat credit expansion as a measure of money stock expansion, which naturally boosts output under the Keynesian framework. On the contrary, the trauma from the Great Recession makes central bankers believe that disorderly credit expansion can be bad for the economy and thus emphasize the issue of debt sustainability ever since. When including data after 2008, many people provide evidence of damage from credit growth. It is interesting to figure out under what circumstances the credit expansion will be good or bad.

The contradictory findings between credit and GDP motivates a multiple-equation setup. The endogeneity among all the macroeconomic variables usually lead to having opposite reduced-form evidence when running a single-equation regression. The success of monetary structural VAR opens a door for distinguishing the opposite-sign causal relationship between macroeconomic variables, for example interest rate and inflation. Hence, the structural VAR setup has been be used in dis-tangling the different links between credit and GDP, and also examine their interaction with monetary policies, as in Brunnermeier et al. (2021). However, there are still not many papers discussing how to use structural VAR models in a cross-country dataset.

This paper tries to estimate a structural VAR to examine different causal channels that lead to comovement between aggregate credit and GDP. By using a balanced panel of 10 countries, we find that, among all the six structural shocks in our model, there is one that pushes aggregate credit up and then later depresses GDP growth, but its effects are small in most countries. All the other five structural shocks still generate positive or zero comovement between these two variables, in line with conventional thoughts. The shocks

[1]Author: Dake Li dakel@princeton.edu, Christopher A. Sims sims@princeton.edu. We received helpful comments from Mikkel Plagborg-Møller, Mark Watson, and numerous seminar and conference participants.

in our model are identified via variation across countries in the relative shock size, with the global trend of shocks taken into account and modeled as the global shock. In our Bayesian estimation framework, we also model the dependence between initial conditions and country constants to avoid the small-T dynamic panel bias.

This paper shows how to properly use a medium-scaled structural VAR model in a cross-country panel dataset. Our dataset covers 10 major economies for 25 years and includes 6 important macroeconomic variables. The model is specified parsimoniously to account for both the heterogeneity across countries and the common trend among them. Besides the country-level fixed effect, one key assumption in our model is the shock size varies across countries. We further split the shocks in each country into two components, i.e. global shocks and country shocks, to capture the global trend in the shock values, together with each country's idiosyncratic movement. In this way, we avoid the curse of dimensionality in the panel dataset, and explain rich dynamics within and across countries. To theoretically identify the model parameters, we utilize identification through heteroskedasticity in the cross-sectional dimension. Unlike Brunnermeier et al. (2021), this paper applies identification through heteroskedasticity when the shock size varies across countries, instead of varying across regimes in a specific country. Because each country has different loadings on the global shock and has different variances in their country shock, the impact matrix in our model can be theoretically point-identified following the scheme in Rigobon (2003).

Following the common routine in monetary structural VAR literature, our model is also estimated in a Bayesian framework. First, we impose conjugate prior on lag coefficients and country constants via dummy observation. Not only do we use the standard Minnesota prior as shrinkage for distant lag coefficients, but we also link the country constants with the initial conditions in each country's time series to avoid the dynamic panel bias in the lag coefficients. Second, we also directly impose dispersed non-conjugate prior on other parameters that we cannot derive the recognizable posterior distributions. Third, we optimize our MCMC algorithm to make posterior draws on a subset of the model parameters. In particular, we integrate out lag coefficients and country constants which have recognizable distribution densities, and then use Metropolis algorithm to draw the rest of the parameters that we cannot analytically integrate out.

In our main results, we analyze the comovement of aggregate credit and GDP growth, and also justify our model specification. We examine among all the six structural shocks in our model, and find that there is one shock coming from the real estate sector, which we label as real-estate shock, that leads to the unusual negative comovement of credit and GDP. However, the average effect of this shock is minimal compared to other sources of variation in credit and GDP. All the other structural shocks still generate the usual positive or zero comovement. This shows that the relationship we observe really depends on which structural shock is leading the variation in a certain period. We also present evidence of huge variation in the shock size across countries, from both the different loadings of global shocks and the different variances of country shocks, which helps to sharply identify our impact matrix. In addition, the global shock that we introduced into our model effectively captures the common trend in the shock terms from this cross-country panel dataset, without too complicated model parameterization.

We additionally run a few robustness checks for our main results. First, we twist the sample of countries and the sample periods to show that our main results are still preserved with another sample in the dataset. In particular, removing US from the sample does not change our main results, indicating it is reasonable to treat US as other countries in this panel VAR model. Second, we try to follow the convention in structural VAR to

recursively define the shocks without using identification through heteroskedasticity, and we largely get the same results. This indicates that identification through heteroskedasticity introduces flexibility in specifying the impact matrix, without generating mysterious results at odds with the conventional methodology.

LITERATURE. Our empirical framework is inspired by Brunnermeier et al. (2021), which establishes a medium-scaled structural VAR on US data to study credit and GDP. That paper explores the different interactions of credit and GDP under different structural shocks, and noticed that the monetary policy can endogenously contribute to the negative comovement. Their model is also identified through heteroskedasticity, with the setup of multiple regimes in the US history. However, that paper only handles the time series data in the US, while our paper tries to extend the setup to the cross-country panel data and handle both the heterogeneity and the common trend across countries.

Numerous work has been proposed to find effects of credit growth in a cross-country panel data. Some old research work favors the traditional idea that credit expansion is good for the economy. Rajan & Zingales (1998) is one example showing the benefit of credit expansion in lowering financing cost of firms. However, the public opinion changed dramatically after 2008. Jordà et al. (2016) relies on a single-equation projection method in a long-run dataset to document the risk from the increased credit level. Mian et al. (2017) studies an unbalanced panel of 30 countries and find evidence for harmful household credit growth across countries. But their paper only sets up a reduced-form VAR model and did not try to model the global trend in the innovation terms. Thus, their conclusion is not exactly a causal interpretation of credit and GDP under different scenarios. Our paper tries to parsimoniously model the rich dynamics across countries and across time, and manage to provide a causal interpretation of comovement between credit and GDP.

The methodology for cross-country panel data has been long studied in the literature. Canova & Ciccarelli (2004) proposes a framework to stack multiple-country variables in a large-scaled VAR model. Because the curse of dimensionality quickly kicks in a large-scaled VAR model, researchers need to make some effort in specifying hierarchical prior to solve the problem. From a different perspective, Pesaran et al. (2004) tries to model the interdependence across countries by running a bunch of country-level VARX models, with the weighted average of foreign variables appearing in the equation as exogenous variables. However, the two step estimation routine of their framework makes it unfeasible to write a internally consistent likelihood formula to analyze the model parameters. In addition, neither of these two papers tries to explore the special issue of structural VAR modeling framework in a cross-country panel setup, i.e. the common trend across countries. We take a stance similar to Stock & Watson (2005) to build the factor structure in the error terms, to characterize the common trend across countries. But instead of only using output data across countries as in their paper, we try to build a model for six variables across countries and try to identify a structural VAR model. Although one can simply follow the conventional tricks to transform reduced-form VAR to structural form, such as running Cholesky decomposition on the variance-covariance matrix of the innovations, our paper tries to utilize the features of this panel data to more flexibly identify the structural parameters, while keeping the model parameterization parsimonious.

This paper also uses the existing tools from the literature of dynamic panel estimation. Since our model has the country constants like the fixed-effect model, and also has a small sample size like any typical macro dataset, the bias presented by Nickell (1981) is a valid concern. One solution, as shown in Sims (2000), is to model dependency of initial conditions with the fixed-effect constants, as the cross-sectional dimension

grows. To model this dependency, one good example is to specify the correlated random effect as Liu (2022), in which the constants are assumed to have a linear relationship with initial conditions.[2] However, in order to easily add dummy observations, we follow the dummy observation representation of the cointegration prior (Sims & Zha, 1998), to model the initial conditions being centered around the implied mean of the model.

Outline. Section 3.2 describes the dataset and the model specification, together with the identification scheme in theory. Section 3.3 explains the steps in our Bayesian estimation procedures. Section 3.4 presents our main results regarding comovement of credit and GDP, variation in the shock size across countries, and benefits of having global shocks. Section 3.5 lists a few robustness checks that we experiment to show the robustness of our main results. Section 3.6 concludes the main contributions of this paper and offers guidance for future research.

3.2 Data and model

This section presents the data and the model that we use in this paper. The details of the data, including the macroeconomic variables and the countries that we select, are listed in Section 3.2.1. We discuss the specification of our structural VAR model, which incorporates global shocks and country shocks, in Section 3.2.2. In terms of model identification, our identification through heteroskedasticity is illustrated in Section 3.2.3.

3.2.1 Data

We use a balanced panel of 10 countries, each with quarterly observations of 6 important macroeconomic variables.

Our dataset incorporates time series across countries in the BIS database and the IMF database. To get a long time series in a balanced panel, we focus on 10 countries with the highest data availability: Australia (AU), Canada (CA), Denmark (DK), United Kingdom (GB), Israel (IL), Japan (JP), Norway (NO), New Zealand (NZ), Sweden (SE), and United States (US). For each country, we gather the data of household credit, policy rate and property price index from the BIS website, and collect the data of GDP, GDP deflator from the IMF website. Moreover, we collect US commodity price index from FRED and adjust the US commodity price by each country's spot exchange rate to get each country's commodity price. In our balanced panel dataset that we finally use, we construct the variables of real property price, real GDP growth, policy rate, GDP deflator, commodity price, household credit in each country from 1995:Q1 to 2019:Q4.[3]

As an illustrative example, the historical paths of household credit in these 10 countries are showed in Figure 3.1. We can see there are huge discrepancies across countries. For example, Japan (JP) had credit contraction after 2000, US had a large credit decline after 2008, but most other countries have a upward-sloping path. The rich heterogeneity in credit paths can help us generalize our results to different country

[2] One can derive the dummy observation representation for correlated random effects, but we avoid using this complicated dummy observation representation in order to make our model specification clear.

[3] When we feed the data into our model, we use policy rate with its decimal numbers, and use all the other variables with its logarithm.

scenarios.

3.2.2 Model specification

Our structural VAR model is specified parsimoniously to count for both the heterogeneity across countries and the global trend in shocks.

Consider country i's variables follow a structural VAR model:

$$Ay_{it} = c_i + B(L)y_{i,t-1} + \nu_{it}, i = 1, 2, \cdots, n, t = 1, 2, \cdots, T, \tag{3.1}$$

where y_{it} is a vector including $m = 6$ variables for country i at time t. All the countries share the slope coefficient $B(L)$ with $p = 8$ lags, but they have different country constants c_i. The shock term ν_{it} also contains $m = 6$ elements, i.e. $\nu_{it} = (\nu_{it}^1, \cdots, \nu_{it}^m)'$.[4] Particularly in the panel setup, we can expect some global trend in the shock term ν_{it}^j across i in each equation j at each time t, and thus the shock term ν_{it}^j can be further modeled with two components as:

$$\nu_{it}^j = \Gamma_i^j f_t^j + \eta_{it}^j, \tag{3.2}$$

where f_t^j is a 1-dim global shock in this factor representation and η_{it}^j is a 1-dim country shock.[5] As in any standard factor model, Γ_i^j is the country i's loading on the global shock f_t^j. We can stack the shocks across multiple equations in country i as:

$$\underbrace{\begin{bmatrix} \nu_{it}^1 \\ \vdots \\ \nu_{it}^m \end{bmatrix}}_{\nu_{it}} = \underbrace{\begin{bmatrix} \Gamma_i^1 & & \\ & \ddots & \\ & & \Gamma_i^m \end{bmatrix}}_{\Gamma_i} \underbrace{\begin{bmatrix} f_t^1 \\ \vdots \\ f_t^m \end{bmatrix}}_{f_t} + \underbrace{\begin{bmatrix} \eta_{it}^1 \\ \vdots \\ \eta_{it}^m \end{bmatrix}}_{\eta_{it}}, \tag{3.3}$$

which clearly illustrates the factor structure in ν_{it}.

For the shock distribution, we assume:

$$\begin{aligned} f_t^j &\overset{\text{iid}}{\sim} \mathcal{N}(0, 1), \forall t, j, \\ \eta_{it}^j &\overset{\text{iid}}{\sim} \mathcal{N}(0, (\Lambda_i^j)^2), \forall i, t, j, \end{aligned} \tag{3.4}$$

which indicates all the shocks in (3.2) are orthogonal to each other, independent across time, and thus have causal interpretation. Only the shocks in (3.2) are called as structural shocks in our model, in which the global shock captures the synchronized trend in ν_{it}^j across i, and the country shock represents the idiosyncratic source of variation. Note that in (3.4), f_t^j is normalized to have standard deviation of 1, and η_{it}^j is normalized to have an average standard deviation of 1 across countries, i.e. $\frac{1}{n}\sum_i \Lambda_i^j = 1$. For simplicity, we also use the

[4] To clearly distinguish the equation index j with the country index i and time index t, this paper always uses a superscript for the equation index j.

[5] To stay in line with the setup of VAR literature, we assume one global shock and one country shock per equation. However, it is possible to extend our setup to include multiple global shocks and one country shock per equation.

Historical Path of Credit

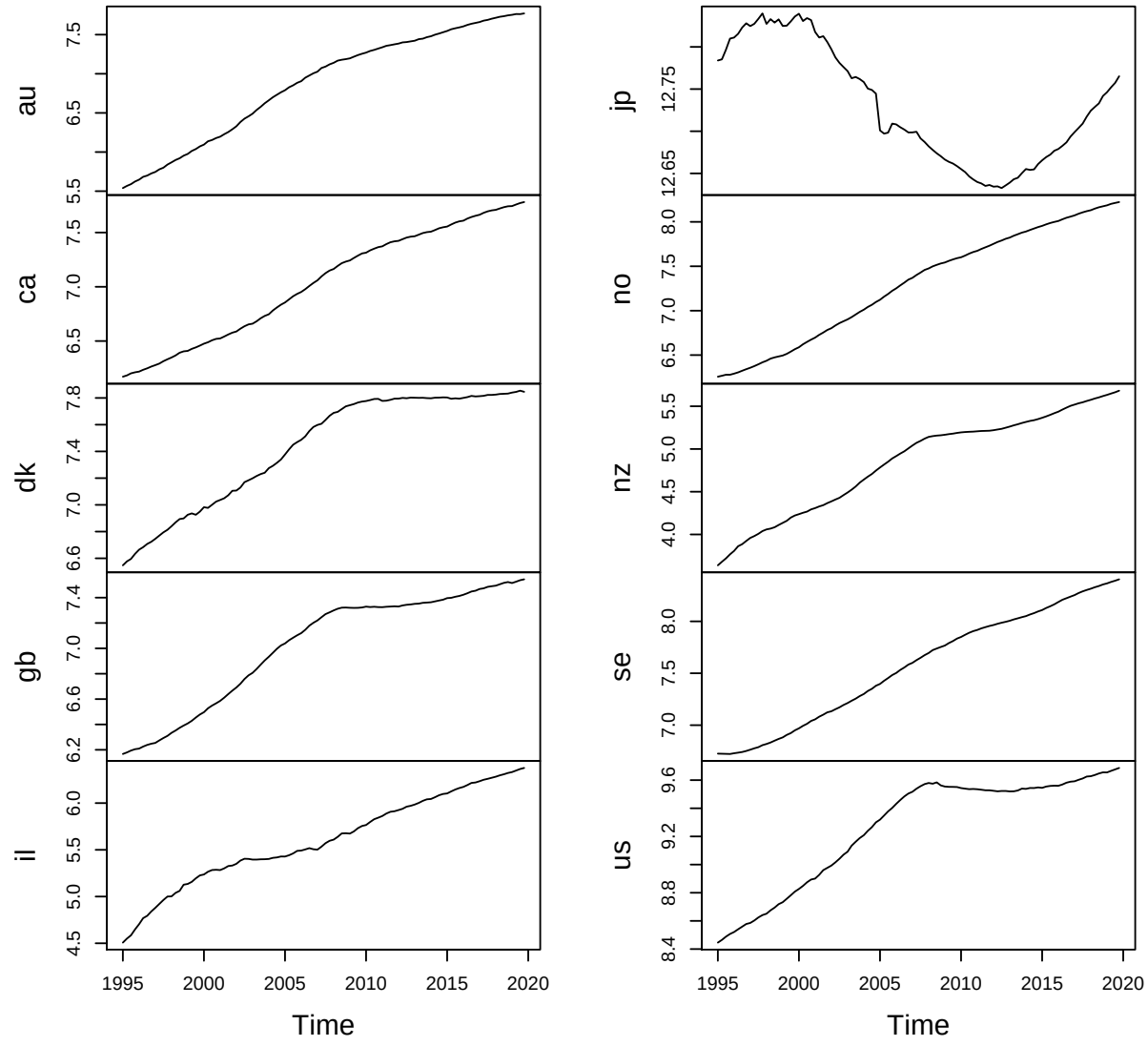


Figure 3.1: We collect the total household credit of each country in its domestic currency (billions). Each credit variable in our model is taken the logarithm of their levels.

aggregate notation $\Lambda_i = \text{diag}(\Lambda_i^1, \cdots, \Lambda_i^m)$.[6]

3.2.3 Identification

The way we identify our structural VAR model is mainly through variation of the shock size across countries, with both global shocks and country shocks taken into account.

The identification of the impact matrix A is the key problem in any structural VAR model, and we identify it in a flexible way via the identification through heteroskedasticity. Unlike imposing recursive ordering and force some elements in A to be zero, we allow all the elements in A to be non-zero. Our identification scheme works when the shock size varies a lot across countries, including both the loaded global shock and the idiosyncratic country shock. By assuming the same impact matrix A in (3.1), the variation of the shock size can help distinguish different structural shocks in different equations of (3.1).

Now we explain how the identification through heteroskedasticity works in our model setup. Consider the reduced-form innovation of y_{it} and define the variance-covariance matrix of reduced-form innovation as:

$$\Sigma_i = \text{Var}(y_{it}|y_{i,(t-\infty):(t-1)}) = \text{Var}(A^{-1}\nu_t) = A^{-1}\,\text{Var}(\nu_{it})A^{-1'}, \tag{3.5}$$

Moreover, we can further derive the variance of ν_{it} from (3.3),

$$\text{Var}(v_{it}) = \text{Var}(\Gamma_i f_t + \eta_{it}) = (\Gamma_i)^2 + (\Lambda_i)^2, \tag{3.6}$$

where $\text{Var}(v_{it})$ clearly has a diagonal structure due to the same structure in Γ_i and Λ_i, and it corresponds to the structural definition of the global shocks and the country shocks.

Finally, we can follow the technique of identification through heteroskedasticity by applying eigenvalue decomposition. where Σ_i is supposed to be identified directly from the data, as one part of the reduced-form specification. After we collect $\{\Sigma_i\}_{i=1}^n$ across countries, we will notice that for any two countries i_1 and i_2:

$$\Sigma_{i_1}^{-1}\Sigma_{i_2} = A'((\Gamma_{i_1})^2 + (\Lambda_{i_1})^2)^{-1}((\Gamma_{i_2})^2 + (\Lambda_{i_2})^2)A^{-1'}, \tag{3.7}$$

which pins down the impact matrix A by computing eigenvalue decomposition.[7]

Once we pin down the elements in A, which equivalent pins down the definition of structural shocks in each equation, the rest of the parameters can also be identified using conventional routines. First, the reduced-form specification together with the identified A can identify the slope coefficient $B(L)$ and the country constants c_i. Second, with the identified A, we can back out the shock term ν_{it} and then identify Γ_i and Λ_i as the factor loading and idiosyncratic variance from standard factor models.

[6] For simplicity, we later also use the aggregate notation Λ, Γ, c to denote the corresponding set of Λ_i, Γ_i, c_i across i.

[7] This identification scheme still pins down up to relabeling of rows, as the literature has noticed. But following Brunnermeier et al. (2021), we impose a dispersed prior on A to nudge the main diagonal of A to be larger than the off diagonal, in order to get around of the relabeling issue.

3.3 Estimation procedure

This section explains our Bayesian estimation procedure in this paper. In Section 3.3.1, we show the technique of using dummy observations as a convenient way to impose prior on autoregressive coefficients and country constants. We also discuss other priors we directly impose on the rest of the parameters that we cannot analytically integrate out in Section 3.3.2. The brief outline of our MCMC algorithm for making posterior draws is listed in Section 3.3.3.

3.3.1 Dummy observations

As a way to apply shrinkage and express prior belief, we use dummy observations to govern the prior of the slope coefficients and the country constants.

The prior on the slope coefficients and the country constants is conditional on A, Γ_i and Λ_i, and thus can be viewed as handling the shrinkage on the reduced-form specification of each equation. Following Sims & Zha (1998), we use three types of dummy observations for each country i:

1. Cointegration: one dummy observation[8]

$$\lambda_1 \cdot A\bar{y}_i = \lambda_1 \cdot c_i + \lambda_1 \cdot \sum_{l=1}^{p} B_{\cdot,\cdot,l}\bar{y}_i + \nu_{i,t_1^*}, \tag{3.8}$$

   where the initial condition denotes $\bar{y}_i = \frac{1}{p}\sum_{l=1}^{p} y_{i,1-l}$, which captures the initial levels of country i's variables.[9] The usage of this cointegration prior can model the dependence between country constants and initial conditions, and can thus address the dynamic panel bias as shown in Sims (2000).[10]

2. Single unit-root: one dummy observation for each variable j

$$\lambda_2 \cdot A(\bar{y}_i \circ e_j) = \lambda_2 \cdot \sum_{l=1}^{p} B_{\cdot,\cdot,l}(\bar{y}_i \circ e_j) + \nu_{i,t_2^*}, \tag{3.9}$$

   where $\circ$ means element-wise product and e_j means a zero vector only with 1 in index j, i.e. we are selecting the j-th element of a vector.

3. Minnesota (random-walk) prior: one dummy observation for each lag l of each variable j

$$\begin{aligned} \lambda_{31} \cdot A(\hat{\sigma} \circ e_j) &= \lambda_{31} \cdot B_{\cdot,\cdot,1}(\hat{\sigma} \circ e_j) + \nu_{i,t_3^*}, \quad \text{if } l = 1, \\ 0 &= \lambda_{31} l_{32}^{\lambda} \cdot B_{\cdot,\cdot,l}(\hat{\sigma} \circ e_j) + \nu_{i,t_3^*}, \quad \text{if } l > 1, \end{aligned} \tag{3.10}$$

   where $\hat{\sigma}$ is a rough estimate of the standard deviation of the reduced-form residuals in the VAR model on y_{it}.

[8] Here one dummy observation means one set of variable values in the multiple-equation system as in (3.1).

[9] We use t^* to denote the auxiliary time index in dummy observations.

[10] Although our dataset has 10 countries and 100 real observations for each country, which is not similar to the usual "big n small T" setup in panel data, the bias from the dynamic panel is still a valid concern given any finite T.

When imposing these dummy observations as the prior, we need to tune the hyper-parameters to control the tightness of the prior. Here we select $\lambda_1 = 0.2/\sqrt{m}$, $\lambda_2 = 0.2/\sqrt{m}$, $\lambda_{31} = 0.1/\sqrt{m}$ and $\lambda_{32} = 1$, to keep the prior reasonably loose and control the magnitude of the residuals in these dummy observations. Since we add these three types of dummy observations for each country i, the division by $\sqrt{m}$ in these hyper-parameters makes sure that we are not going to impose a tighter and tighter prior as the number of countries m grows.

In this way, we construct the above-mentioned set of dummy observations for each country and model the country constants with their initial conditions. The dummy observation of cointegration penalizes the country constant if it implies a steady state too far away from the initial conditions, since those cases occur when the huge bias of the country constants kicks in to mechanically use the mean-reversion trend to fit the persistent time series. This is a special concern in dynamic panel data, since we have a lot of country constants and the time span is not extremely long.

3.3.2 Other priors on parameters

In addition to dummy observations, we also impose other priors directly on parameters A, Λ_i, Γ_i.

Since it is difficult to derive a recognizable posterior distribution for A, Λ_i, Γ_i in (3.1) - (3.2), we directly use non-conjugate prior for these parameters. For each element in A, we use a Gaussian prior:

$$p(A_{jk}) = \mathcal{N}(1\{j = k\} \cdot 100, 200^2), \tag{3.11}$$

where the dispersed prior slightly push the diagonal of A to the positive region and to be slightly larger than the off diagonal, in order to get around the well-known relabeling issue of A. For Λ_i, given the normalization of having its average equal to 1 across i, we impose a scaled Dirichlet prior on Λ_i in each equation j:

$$p(\Lambda_1^j, \cdots, \Lambda_n^j) \propto \text{Dir}(2, \cdots, 2). \tag{3.12}$$

Finally for Γ_i, we use a truncated Gaussian prior. However, to avoid introducing sharp discontinuity in the posterior density, we first use a untruncated Gaussian prior for Γ in each equation j to make posterior draws, and then later map the draws to the target region where the average of Γ_i in each equation j are normalized to be positive:

$$p(\Gamma_1^j, \cdots, \Gamma_n^j) \propto \mathcal{N}(0, 1), \quad \text{if } \frac{1}{n}\sum_{i=1}^{n} \Gamma_i^j > 0 \tag{3.13}$$

Here the prior on Γ indicates that ex ante we expect the global shock component is as large as the country shock component for an average country. All the prior has been set to be relatively dispersed to cover the possible range of parameter values.

3.3.3 MCMC algorithm

We set up a MCMC algorithm, which fully utilizes the structure of this model, to make posterior draws. There are two parts in evaluating the posterior distribution for our MCMC algorithm:

1. Integrate out the parameters with recognizable posterior distribution, i.e. integrating c_i and $B(L)$ given A, Λ_i and Γ_i,[11]

$$p(Y|A,\Lambda,\Gamma) = \int_{c,B} p(c,B(L)|A,\Lambda,\Gamma)\cdot p(Y|c,B(L),A,\Lambda,\Gamma)\cdot dc\cdot dB, \tag{3.14}$$

   Since it is to verify the posterior distribution of c_i and $B(L)$ given other parameters is Gaussian, we can easily integrate out c_i and $B(L)$ to get the marginal likelihood for A,Λ,Γ.[12]

2. Directly evaluate posterior for parameters that cannot be integrated out,

$$p(A,\Lambda,\Gamma|Y) \propto p(A)p(\Lambda)p(\Gamma)p(Y|A,\Lambda,\Gamma), \tag{3.15}$$

   where $p(Y|A,\Lambda,\Gamma)$ is obtained from (3.14).

To make the MCMC algorithm computationally efficient, we apply Metropolis method on (3.15) to only draw A,Λ,Γ. We execute a long MCMC series for these three parameters, and then later draw c and $B(L)$ given these three parameters from a Gaussian posterior when analyzing impulse responses and shock sizes. Our MCMC algorithm generates 100,000 draws, from which we later thin to keep only 100 draws for the inference results. Further details are discussed in Section 3.A.

3.4 Main results

This section presents the main results from estimating our structural VAR model. The comovement between credit and GDP, under scenarios of different structural shocks, are described in Section 3.4.1. We also compare the shock size across different countries, in terms of their loaded global shocks and their idiosyncratic country shocks, as in Section 3.4.2. Furthermore, we show the benefits of including global shocks into our model in Section 3.4.3.

3.4.1 Comovement of credit and GDP

We analyze the comovement of credit and GDP, with the occurrence of different structural shocks, to unveil the various causal channels underlying the documented relationship.

Since we identify the impact matrix through heteroskedasticity, we allow all the elements of our impact matrix to be non-zero. Our estimate of A at the posterior peak is presented in Table 3.1. The impact matrix has small entries off-diagonal, capturing the small correlations across variables. Its main-diagonal entries are much larger, and thus we label the shocks in line with the variable names.

We document the impulse responses of credit and GDP, together with other included macroeconomic variables, as in Figure 3.2. Since we have six equations, we will have six types of structural shocks, where each type has its global shock component and its country shock component. Here we examine the responses when an average country has a one-standard-deviation move for each type of structural shocks, where its standard

[11] For simplicity, we also use the aggregate notation of Λ,Γ,c to denote the set of Λ_i,Γ_i,c_i across i.

[12] We denote all the real observations as Y, and all the dummy observations as Y^* in our illustration.

Posterior Modal Estimate of Impact Matrix

	real property	real GDP	policy rate	commodity price	GDP deflator	household credit
real-estate shock	92.94	-1.11	0.08	2.57	0.46	-1.00
output shock	-1.99	105.35	-0.93	0.37	0.79	-0.50
monetary shock	-0.58	-1.30	105.54	0.14	-0.40	-0.44
commodity shock	3.83	0.81	0.26	31.94	-2.95	-0.09
price shock	2.88	2.14	-0.33	-1.26	103.43	-0.47
credit shock	-3.47	-1.08	-0.32	0.36	-1.25	105.05

Table 3.1: Each row represents one equation in our model.

deviation includes the variation of both its global shock component and its country shock component, in order to demonstrate which type of shocks explains most of the variation for most countries.

Across all the six structural shocks, we find the shock from the real estate sector largely drives the negative comovement of credit and GDP. Here, since the first shock in Figure 3.2 pushes up the property price a lot, we can thus label the first shock as the real-estate shock. This real estate shock generates an increasing trend in credit, but the response of GDP is hump-shaped, with a negative GDP growth rate after 1 year. In addition, we note that the policy rate endogenously rises shortly after this real-estate shock, which contributes to this negative relationship. This seems the only exceptional pattern across all the impulse responses, because all the other five structural shocks are either generating positive comovement or no comovement between credit and GDP, which matches the convectional opinions before 2008. For example, the monetary policy shock (the third shock in Figure 3.2) will lower GDP and credit at the same time, which is the typical effect of monetary tightening.

The exceptional negative relationship explains the puzzle in contradictory opinions about credit growth. First, the shocks coming from different sectors will lead to different comovement, and the real estate shock plays an important role whenever we observe bad effects of credit growth. Second, the magnitude of variation caused by this real estate shock is small, compared to other sources of variation in credit and GDP, which explains why we did not collect much evidence on bad credit growth before 2008. Third, it is noteworthy that the monetary policy rate endogenously responds to the real estate shock, which drives down GDP after 1 year. Any analysis omitting this endogenous interaction might lead to puzzling conclusions.

3.4.2 Size of global shocks and country shocks

We show that there is huge heterogeneity in the shock size across countries, in terms of both the global shock component and the country shock component.

First, we present the size of idiosyncratic country shocks, i.e. the posterior peak value of Λ, as in Table 3.2. We notice that the second shock, i.e.output shock, has similar sizes across countries. On the contrary, the third shock, i.e. monetary policy shock, is more than 100 times larger in some countries than the others.

Now we turn to each country's loading on the global shock, i.e. the posterior peak value of Γ, as shown in

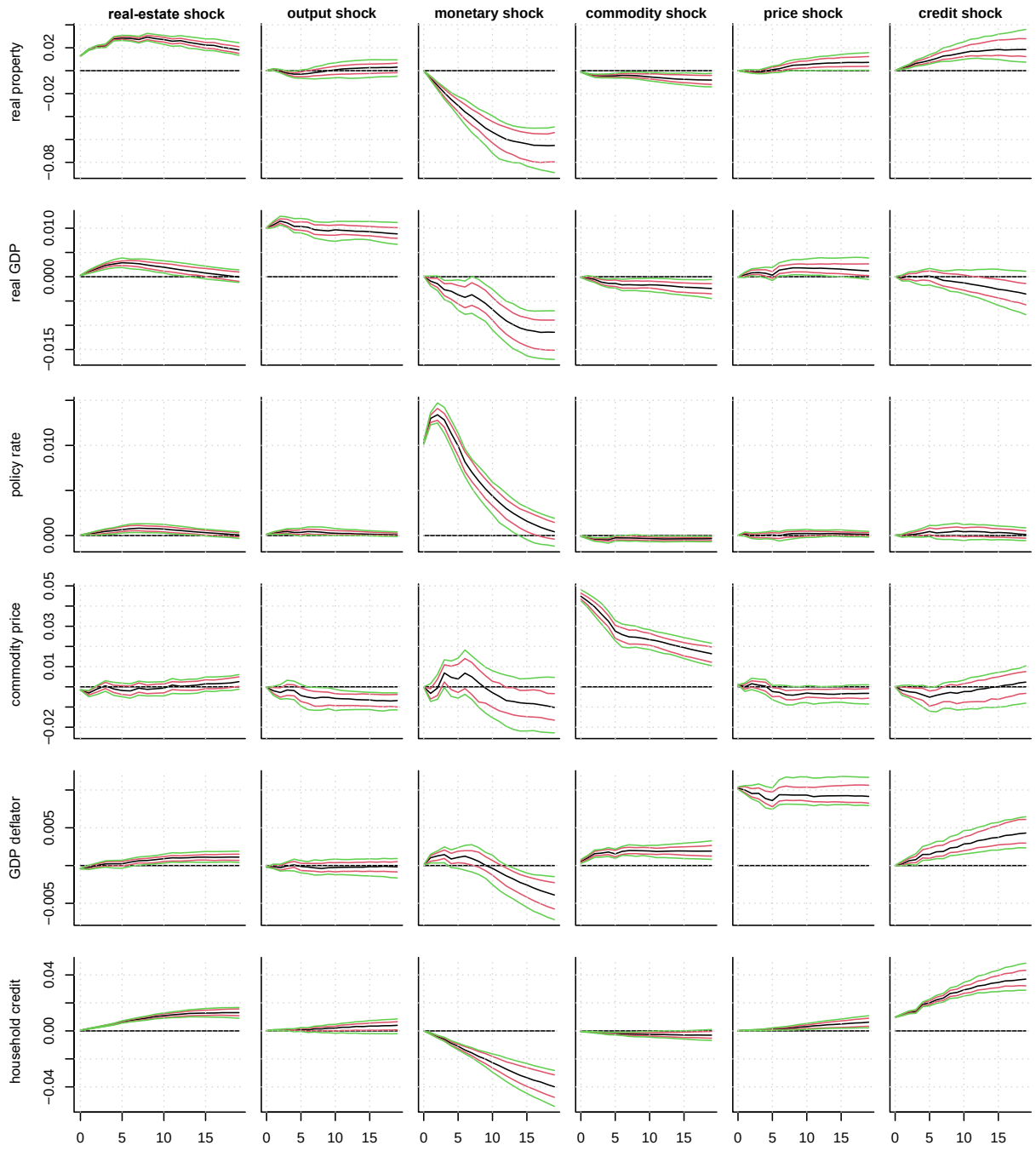


Figure 3.2: The 68% and 90% credible bands are plotted in red and in green. We feed in a one-std shock (including both variation of global and country shocks) for each type of structural shock in an average country, which equals to the average shock size across countries.

Posterior Modal Estimate of Standard Deviation in Country Shock

	AU	CA	DK	GB	IL	JP	NO	NZ	SE	US
real-estate shock	1.22	1.11	1.14	0.03	1.14	0.83	1.54	1.01	1.21	0.76
output shock	0.54	0.48	0.82	0.35	0.88	1.09	3.59	1.02	0.76	0.48
monetary shock	0.24	0.25	0.28	0.19	7.39	0.07	0.45	0.60	0.21	0.32
commodity shock	1.30	0.90	0.78	1.00	0.93	2.02	0.94	1.38	0.09	0.66
price shock	0.83	0.79	0.54	0.75	1.07	0.34	4.00	1.05	0.50	0.12
credit shock	0.57	0.49	1.01	0.52	5.04	0.73	0.25	0.47	0.25	0.67

Table 3.2: Each row is normalized to have average of 1, in order to pin down the scale of each structural equation.

Posterior Modal Estimate of Loading on Global Shock

	AU	CA	DK	GB	IL	JP	NO	NZ	SE	US
real-estate shock	0.37	0.06	0.40	1.53	0.11	-0.03	0.28	-0.13	0.41	0.14
output shock	-0.04	0.23	0.28	0.19	0.39	0.44	0.03	-0.04	0.44	0.22
monetary shock	0.26	0.23	0.17	0.33	0.23	0.02	0.29	0.19	0.30	0.24
commodity shock	0.99	0.49	1.28	0.73	0.51	0.42	1.15	1.03	1.50	0.01
price shock	0.41	0.44	0.21	0.05	-0.21	-0.25	0.73	0.12	0.00	0.06
credit shock	0.11	-0.12	-0.36	-0.01	0.17	0.02	0.12	0.19	-0.07	0.06

Table 3.3: Each row is normalized to have average above 0, in order to uniquely pin down the sign of global shocks.

Table 3.3. First, most countries have positive loadings, which means they follow the global trend in the same direction. This coincides with our usual observations that the world economy comoves together. Second, a certain country can follow the global trend in the shock, and go against the global trend in the other shock. For example, Canada (CA) has big positive loadings on shock 4 (global commodity shock) and shock 5 (global price shock), but big negative loading on shock 6 (global credit shock). Third, across all the different shocks, shock 4 (global commodity shock) has the largest cross-country linkages. Since shock 4 (commodity shock) is mainly from the commodity market, it makes sense to have strong synergy in this shock, given the integration of the global commodity market and the way in which we construct each country's commodity price index.

We can now compare the variation of shock sizes in each country's global shock component and country shock component. Figure 3.3 compares the standard deviation of the global shock component and country shock component across countries. We pick the real-estate shock and monetary policy shock here as two illustrative examples (See Section 3.B for this comparison in other shocks). First, we notice huge variation in contribution of global shocks vs. country shocks. For example, GB's real-estate shocks relies heavily on the global trend, while IL's monetary policy shocks are mostly idiosyncratic. Second, the importance of the global trend differs across types of shocks. For most countries, the global trend plays a limited role for

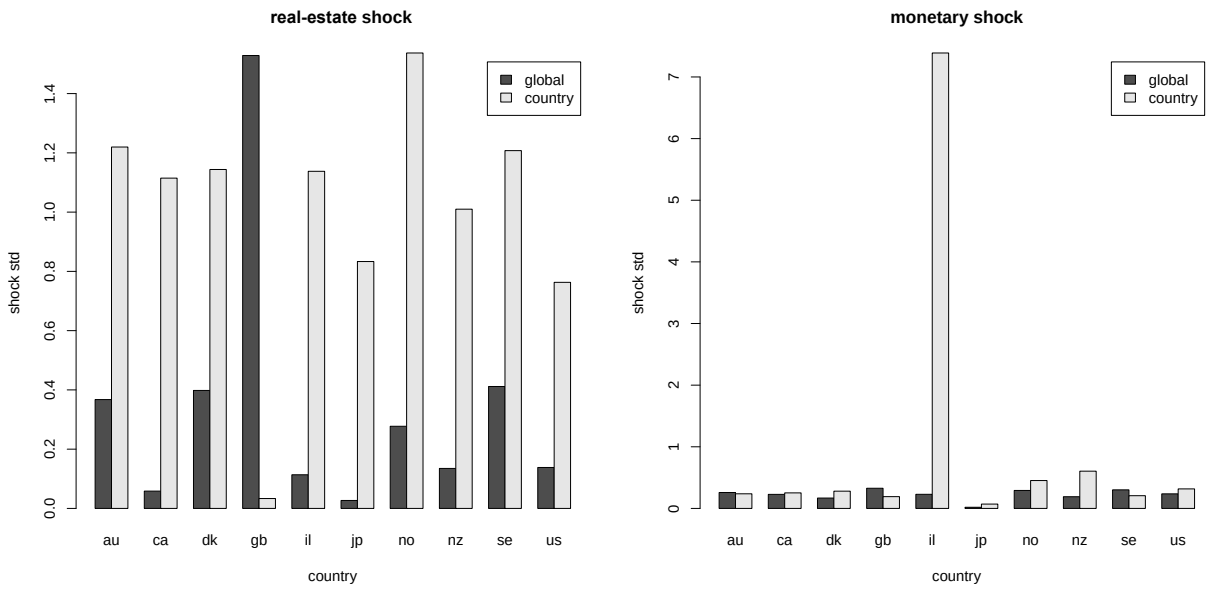


Figure 3.3: This bar plot shows the standard deviation of the global shock component and country shock component across countries. The standard deviation of global shock component is the absolute value of Γ. The standard deviation of country shock component, i.e. Λ, is normalized to have an average of 1 across countries.

real-estate shocks, while the global trend imposes a stronger synergy in monetary policy shocks. Third, there is no clear pattern showing the complementary or substitution between the global component and country component, because we do not see the standard deviation of these two components always moving in the same or opposite direction.

3.4.3 Comparison with and without global shocks

Including global shocks brings many benefits into our model, including describing the global trend, identifying the model, and improving the model fit.

We document the time series of implied global shock, as in Figure 3.4. Here we consider both the uncertainty from the model parameters and from the latent shocks. In our specification, the global shock can be interpreted as contemporaneous spillover of shocks across countries. For example, central banks will decide the policy rate next period and introduce some unexpected surprise across time. But when they are in the process of determining the policy rate for next period, they might have meetings with other central banks to communicate, and thus create some global trend in the monetary policy shock next period, which is captured by the global monetary policy shock listed in the graph. After taking out the global trend, the remaining variation in each country's monetary policy is supposed to be idiosyncratic. We plot the country shocks of monetary policy as an example in Figure 3.5. Across countries, we no longer see dramatic dips in the country shocks of monetary policy around 2008, because the common dip is interpreted as a global shock, and the idiosyncratic deviation from this common dip will show up as the country shock. Thus, most of the common trend across countries has been absorbed into the global shock component, leaving the country

Model fit with / without global shocks

	log of marginal data density
with global shocks	16379.7
without global shocks	16377.2

Table 3.4: We report the logarithm of marginal data density, computed from bridge sampling with a Gaussian proposal density around the posterior peak, using 100 posterior draws.

shock component closer to being orthogonal.[13]

Introducing the global shock into our model has multiple benefits. First, it explicitly models the global component and country component in the shock terms, and effectively captures the global trend. This preserves the causal interpretation in our structural model, because we can now discuss the response of either a global shock or a country shock, without worrying the synergy of shocks across countries. Second, the global shock brings another layer of shock size variation across countries. As we have seen in Section 3.4.2, the loadings on the global shock in each country vary dramatically, which can help the identification scheme to work in addition to the variations in sizes of country shocks, as we can rely on both the change of Γ_i and Λ_i in (3.7). Third, the goodness of model fit mildly increases after the usage of global shocks, indicating it is crucial to specify the right cross-country correlation structure to fit the data. Table 3.4 lists the logarithm of marginal data density using models with and without global shocks.[14] Although using the global shocks remarkably increases the likelihood[15] and easily rejects the null in a frequentist likelihood-ratio test, our Bayesian analysis gives a more conservative conclusion, showing there is only mild increase in the model fit, especially in the case of testing a large amount of free parameters, i.e. the 60 parameters of Γ.[16]

3.5 Robustness checks

This section implements a few robustness checks on our data and model. In Section 3.5.1, we change the countries and the periods in our sample to show our main results are still robust. In Section 3.5.2, we use the conventional identification scheme, by recursively defining shocks and ignoring global trends, to show that the impulse response results are also robust.

3.5.1 Alternative sample

To test the robustness of our main results, we change the sample by altering the selection of countries and the sample period. First, we allow shorter time series to expand the dataset to 13 countries from 2000:Q1

[13] Here we cannot expect the orthogonality to be perfect, because the model specification is still parsimonious and cannot fully interpret every correlation across countries. As shown in Section 3.C, if we impute the country shocks of monetary policy at the posterior peak and compute the correlation matrix, we can still find a few big correlations. However, compared to the model without global shocks, our model has done a good job in tackling down most of the big correlations.

[14] For the model without global shocks, we simply force $\Gamma = 0$, and thus the model only has idiosyncratic country shocks, i.e. $\nu_{it}^j = \eta_{it}^j$ as in standard panel structural VAR models.

[15] The log-likelihood increases by more than 300, which is much larger than $\chi^2_{60,0.95}$.

[16] Although the increase of log marginal data density by 2.5 would imply the model with global shocks is 12 times more likely, the rule of thumb in Bayesian high-dim model comparison would view this as an mild increase, given we are testing two distinct models, with no other models in between.

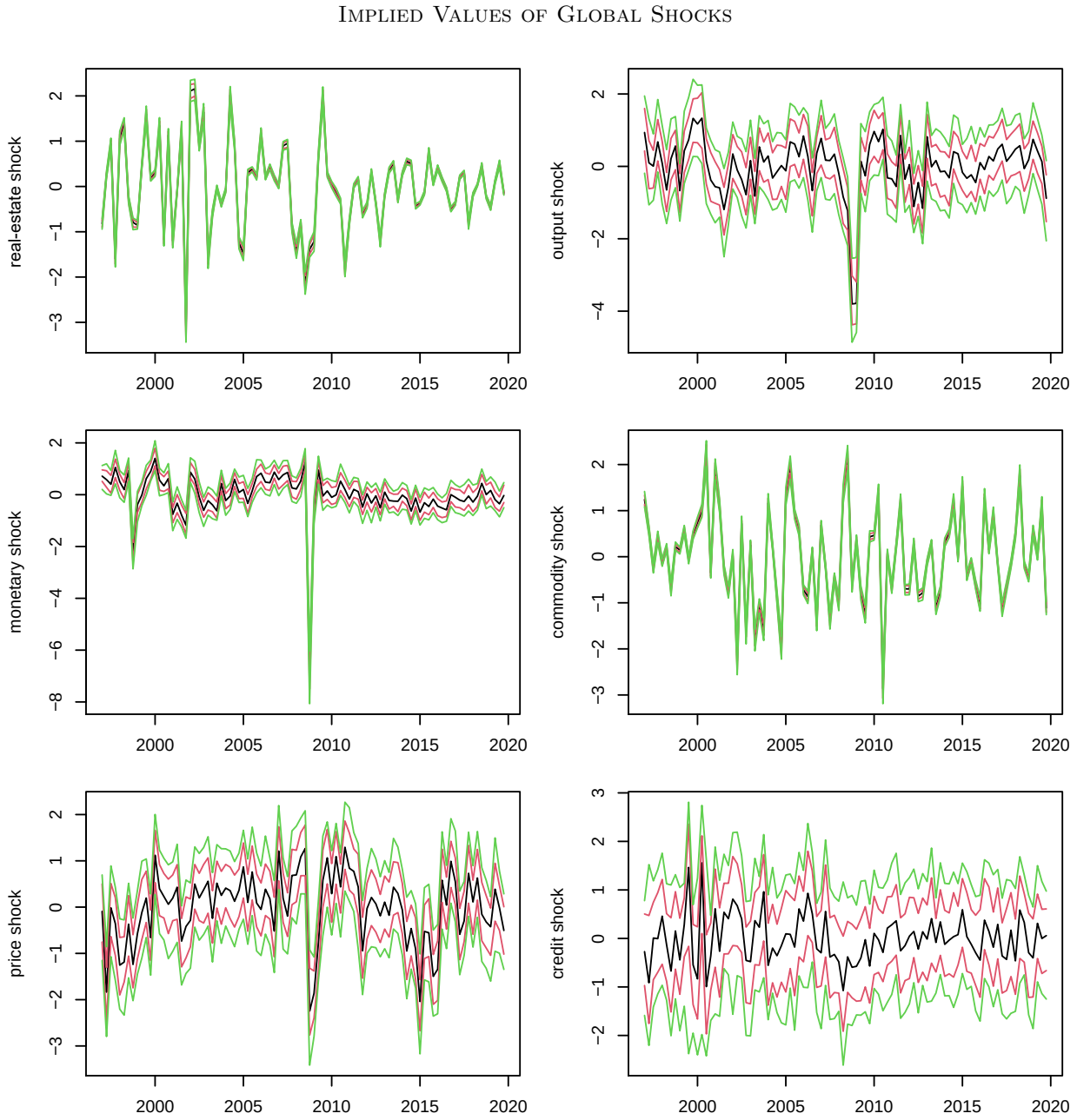


Figure 3.4: The 68% and 90% credible bands are plotted in red and in green. Both the uncertainty from the model parameters and the latent shock values are incorporated. Each global shock is normalized to have mean 0 and standard deviation 1.

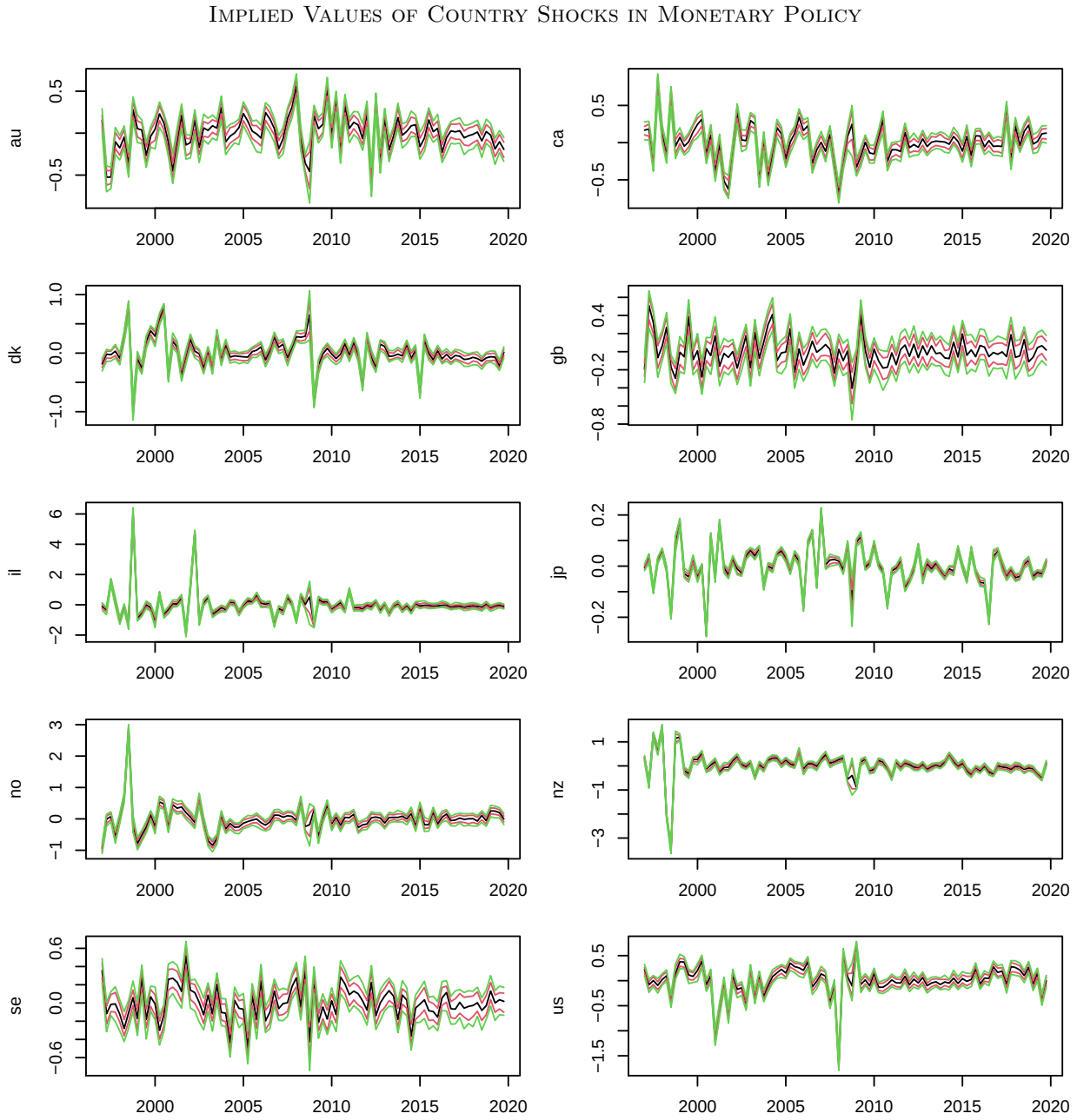


Figure 3.5: The 68% and 90% credible bands are plotted in red and in green. Both the uncertainty from the model parameters and the latent shock values are incorporated. Each country shock is normalized to have mean 0 and average standard deviation 1 across countries.

to 2019:Q4. Our main results are preserved. Second, we remove US in our baseline specification, and re-estimate the model, which generate similar results. This means it is feasible to treat US the same as other countries, as long as we have modeled the cross-country heterogeneity and the global trend in our panel data.

3.5.2 Conventional identification scheme

To show that our identification scheme gives results in line with conventional identification scheme, we re-estimate our model without global shocks and define structural shocks recursively. The implementation simply treats (3.1) as a standard panel structural VAR model with country fixed effects, and estimate the pooled variance-covariance of ν_{it}. Then we use Cholesky decomposition to back out the impact matrix and examine the impulse response. The response largely coincides with our main results, indicating our identification scheme in the main results is allowing extra flexibility while not generating anything completely mysterious.

3.6 Conclusion and directions for future research

We estimate a structural VAR on a balanced panel of 10 countries to examine different causal channels that lead to comovement between aggregate credit and GDP. The model is specified to account for heterogeneity across countries and to capture the global trend. We show this model can be identified through heteroskedasticity, with the variation in the size of loaded global shocks and country shocks. Among all the six structural shocks in our model, we find that there is one from the real estate sector, that pushes aggregate credit up and then later depresses GDP growth, but its effects are small in most countries. All the other five structural shocks still generate the positive or zero comovement between aggregate credit and GDP. The results also imply huge variation in shock sizes across countries. The inclusion of the global shocks effectively captures the global trend of shocks across countries.

There are some interesting extensions of our model specification. First, researchers can specify both the shock size variation across countries and across time. This is going to combine our paper with (Brunnermeier et al., 2021), and use the cross section and time series information to identify the impact matrix. In addition, this type of models will have rich dynamics that can characterize the extremely large shock values around 2008 and even after the pandemic. Second, one can extend our specification by introducing dynamic structures of global shocks. The example of 2008 Financial Crisis shows that the pace of rates cut will largely be synchronized across countries, but might have some time mismatches around 1 quarter. Having lagged terms of global shocks appearing in our model will answer the question that how soon each country follow the global trend, for example, in the common rates cut. Third, one can try some sub-models between our model with global shocks and the model without global shocks. Our analysis shows that adding global shocks for each structural equation brings in too many free parameters, which is not favored using Bayesian information criteria. Simply adding global shocks for equations with large cross-cournntry correlation will be promising.

Appendix 3.A Details on MCMC algorithm

The details of our MCMC algorithm are listed here.

We first optimize the MCMC algorithm by integrating out c and $B(L)$ in the likelihood. Since we use dummy observations to express prior on c and $B(L)$, we need to integrate the likelihood twice to get the correct posterior. Denote the real observations as Y and the dummy observations as Y^*. In addition, the prior that the dummy observations generates is

$$p(c, B(L)|A, \Lambda, \Gamma) = \frac{p(Y^*|c, B(L), A, \Lambda, \Gamma)}{\int_{c,B} p(Y^*|c, B(L), A, \Lambda, \Gamma) dc \cdot dB}. \tag{3.A.1}$$

Thus the likelihood with c and $B(L)$ integrated out both in the real and dummy observations is

$$\begin{aligned} p(Y|A, \Lambda, \Gamma) &= \int_{c,B} p(c, B(L)|A, \Lambda, \Gamma) \cdot p(Y|c, B(L), A, \Lambda, \Gamma) dc \cdot dB \\ &= \int_{c,B} \frac{p(Y^*|c, B(L), A, \Lambda, \Gamma)}{\int_{c,B} p(Y^*|c, B(L), A, \Lambda, \Gamma) dc \cdot dB} p(Y|c, B(L), A, \Lambda, \Gamma) dc \cdot dB \\ &= \frac{\int_{c,B} p(Y, Y^*|c, B(L), A, \Lambda, \Gamma) dc \cdot dB}{\int_{c,B} p(Y^*|c, B(L), A, \Lambda, \Gamma) dc \cdot dB} \end{aligned} \tag{3.A.2}$$

After integrating out c and $B(L)$, we are left with parameters, i.e. A, Λ, Γ, which cannot easily integrate out, and hence we turn to Metropolis algorithm to directly make posterior draws on this smaller set of parameters. We run a pilot series of 10,000 Metropolis draws from the posterior peak, and update the variance-covariance matrix in our Gaussian proposal density. Then we start a formal series of 100,000 Metropolis draws for our main results. We check the convergence of our MCMC draws by using the traceplot of our posterior density. When computing impulse responses or implied shock values, our routine is taking a posterior draw of A, Λ, Γ and then analytically draw c and B given the value of A, Λ, Γ.

Appendix 3.B Shock size variation in global and country components

Here we list out the comparison of the standard deviation in the global shock component and the country shock component, similar to Section 3.4.2

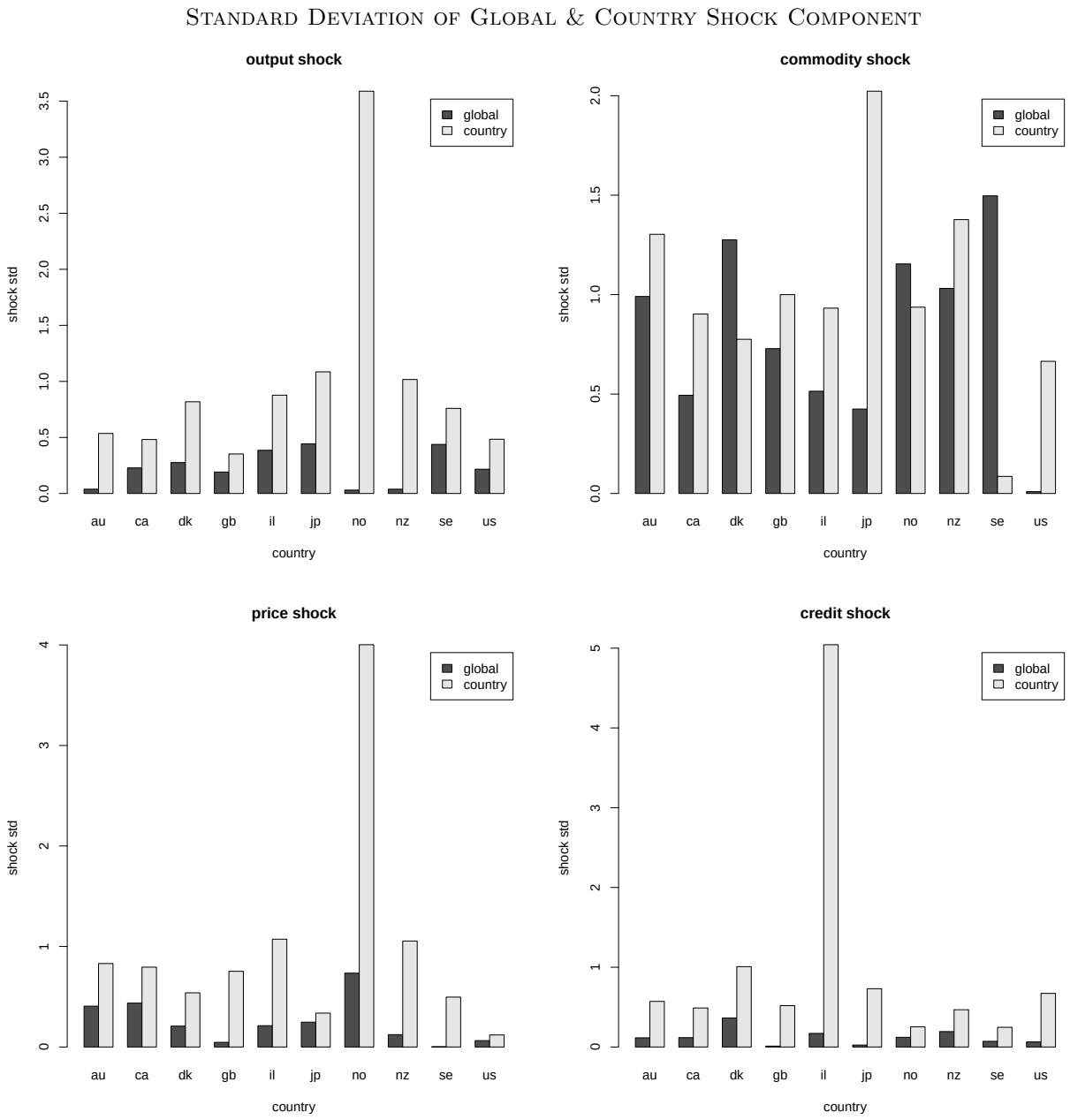


Figure 3.B.1: This bar plot shows the standard deviation of the global shock component and country shock component across countries. The standard deviation of global shock component is the absolute value of Γ. The standard deviation of country shock component, i.e. Λ, is normalized to have an average of 1 across countries.

Appendix 3.C Cross correlation in monetary policy shocks

Here we impute the country shock values at the posterior peak, and compute the sample correlation matrix. We then compare the correlation matrix using the model with global shocks and the model without global shocks, to check how many big correlation entries have been absorbed by the global shocks.

Cross correlation in country shocks in monetary policy

(a) Model with global shocks

	AU	CA	DK	GB	IL	JP	NO	NZ	SE	US
AU	1.00	-0.22	-0.15	-0.31	0.13	0.15	-0.14	0.15	-0.09	-0.05
CA	-0.22	1.00	0.10	-0.26	0.15	-0.11	0.04	-0.11	-0.27	0.23
DK	-0.15	0.10	1.00	-0.11	-0.24	-0.36	0.34	-0.25	-0.12	-0.21
GB	-0.31	-0.26	-0.11	1.00	-0.18	0.03	-0.08	-0.01	-0.33	-0.17
IL	0.13	0.15	-0.24	-0.18	1.00	0.11	-0.13	0.30	-0.04	0.11
JP	0.15	-0.11	-0.36	0.03	0.11	1.00	-0.37	0.29	0.01	0.10
NO	-0.14	0.04	0.34	-0.08	-0.13	-0.37	1.00	-0.57	-0.07	-0.19
NZ	0.15	-0.11	-0.25	-0.01	0.30	0.29	-0.57	1.00	0.04	0.09
SE	-0.09	-0.27	-0.12	-0.33	-0.04	0.01	-0.07	0.04	1.00	-0.29
US	-0.05	0.23	-0.21	-0.17	0.11	0.10	-0.19	0.09	-0.29	1.00

(b) Model without global shocks

	AU	CA	DK	GB	IL	JP	NO	NZ	SE	US
AU	1.00	0.40	0.29	0.59	0.17	0.18	0.30	0.32	0.58	0.46
CA	0.40	1.00	0.34	0.50	0.14	0.06	0.39	0.09	0.43	0.53
DK	0.29	0.34	1.00	0.41	-0.16	-0.12	0.46	-0.05	0.36	0.17
GB	0.59	0.50	0.41	1.00	-0.02	0.13	0.40	0.26	0.68	0.49
IL	0.17	0.14	-0.16	-0.02	1.00	0.08	-0.06	0.32	0.09	0.13
JP	0.18	0.06	-0.12	0.13	0.08	1.00	-0.06	0.22	0.16	0.18
NO	0.30	0.39	0.46	0.40	-0.06	-0.06	1.00	-0.28	0.38	0.17
NZ	0.32	0.09	-0.05	0.26	0.32	0.22	-0.28	1.00	0.31	0.28
SE	0.58	0.43	0.36	0.68	0.09	0.16	0.38	0.31	1.00	0.38
US	0.46	0.53	0.17	0.49	0.13	0.18	0.17	0.28	0.38	1.00

Table 3.C.1: Both tables show the sample correlation matrix of implied country monetary policy shocks at the posterior peak. The first one uses the model with global shocks, while the second uses the model without global shocks.

References

Brunnermeier, M., Palia, D., Sastry, K. A., & Sims, C. A. (2021). Feedbacks: financial markets and economic activity. *American Economic Review*, *111*(6), 1845–79.

Canova, F. & Ciccarelli, M. (2004). Forecasting and turning point predictions in a bayesian panel var model. *Journal of Econometrics*, *120*(2), 327–359.

Jordà, Ò., Schularick, M., & Taylor, A. M. (2016). The great mortgaging: housing finance, crises and business cycles. *Economic Policy*, *31*(85), 107–152.

Liu, L. (2022). Density forecasts in panel data models: A semiparametric bayesian perspective. *Journal of Business & Economic Statistics*, *0*(0), 1–15.

Mian, A., Sufi, A., & Verner, E. (2017). Household debt and business cycles worldwide. *Quarterly Journal of Economics*, *132*(4), 1755–1817.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*(6), 1417–1426.

Pesaran, M. H., Schuermann, T., & Weiner, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, *22*(2), 129–162.

Rajan, R. & Zingales, L. (1998). Financial development and growth. *American Economic Review*, *88*(3), 559–586.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, *85*(4), 777–792.

Sims, C. A. (2000). Using a likelihood perspective to sharpen econometric discourse: Three examples. *Journal of Econometrics*, *95*(2), 443–462.

Sims, C. A. & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, *39*(4), 949–968.

Stock, J. H. & Watson, M. W. (2005). Understanding changes in international business cycle dynamics. *Journal of the European Economic Association*, *3*(5), 968–1006.