Essays on the Effects of Macroeconomic Shocks

Matthew Read

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Economics

University College London

May 22, 2022

I, Matthew Read, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Date: 14 October, 2021

Abstract

This thesis explores the effects of macroeconomic shocks in different contexts. In Chapter 2, I estimate the effects of monetary policy shocks on firm entry and exit rates, and develop a New Keynesian firm-dynamics model that qualitatively matches the empirical responses. I find that accounting for responses along the entry and exit margins does not appreciably change the predictions of the standard (representative-agent) New Keynesian model. Chapter 3 develops algorithms to facilitate Bayesian inference in structural vector autoregressions (SVARs) that have been identified using sign and zero restrictions. Chapter 4 reviews the literature on robust Bayesian inference as a tool for global sensitivity analysis and statistical decision-making under ambiguity. It also presents a self-contained discussion of three different approaches to conducting robust Bayesian inference in set-identified SVARs, along with an empirical example. Chapter 5 explains how to conduct robust (multiple-prior) Bayesian inference in set-identified proxy SVARs, and explores the frequentist properties of this approach when the proxies are only weakly correlated with the structural shocks of interest. It illustrates the approach by estimating the effects of shocks to average income tax rates in the United States. Chapter 6 explores issues related to identification and inference in SVARs under narrative restrictions, which are restrictions on functions of the structural shocks in particular periods.

Impact Statement

This dissertation makes several contributions to the existing literature and contains material that should be useful for policymakers seeking to estimate the effects of macroeconomic policy interventions.

Chapter 2 presents new empirical evidence suggesting that the margins of firm entry and exit may contribute substantially to changes in aggregate employment following a monetary policy shock. Although a standard New Keynesian model augmented with heterogeneous firms and endogenous entry and exit can qualitatively match the responses of entry and exit to a monetary policy shock, the model does not deliver predictions that differ meaningfully from the benchmark New Keynesian model. These results suggest that incorporating the entry and exit of heterogeneous firms into otherwise standard New Keynesian models – such as those used in central banks – is unlikely to yield meaningfully different predictions about the effects of monetary policy, unless the entry and exit margins interact with firm-level frictions.

Chapter 3 proposes algorithms that can be used to facilitate Bayesian inference in SVARs that are set-identified with a mixture of sign and zero restrictions. I show that the algorithms are much more computationally efficient than existing alternatives when there are a large number of sign restrictions that considerably truncate the identified set given the zero restrictions. These algorithms should therefore facilitate the use of rich sets of sign restrictions alongside zero restrictions in empirical settings.

Chapter 4 presents a review of the literature on robust Bayesian inference as it applies to the field of econometrics. This chapter also provides a detailed overview of three different approaches to conducting robust Bayesian inference in set-identified SVARs, including details on numerical implementation and an empirical application. This work should provide a useful reference for practitioners wishing to eliminate or quantify the sensitivity of posterior inference to the choice of prior in set-identified models.

Chapter 5 proposes an approach for conducting prior-robust Bayesian inference in proxy SVARs that are set-identified. The approach is likely to be particularly useful when there are multiple proxies for multiple structural shocks. An important setting where this issue arises is the estimation of the effects of changes in different tax rates. Importantly, the approach allows practitioners to relax potentially controversial point-identifying restrictions (possibly replacing them with arguably weaker sign restrictions) without posterior inference being driven by an unrevisable component of the prior.

Chapter 6 explores issues related to identification and inference in SVARs under ‘narrative restrictions’, which are restrictions on functions of the structural shocks in specific periods. These restrictions are being used increasingly to sharpen identification in SVARs, but they have been subject to little formal analysis. This chapter shows that nonstandard features of the restrictions mean that standard frameworks for thinking about identification do not apply under these restrictions, and existing approaches to Bayesian inference possess undesirable properties. An alternative robust Bayesian approach to inference addresses these issues and is valid from a frequentist perspective. Consequently, the approach should appeal to both Bayesians and frequentists seeking to impose this class of restrictions in empirical analysis.

Acknowledgements

I sincerely thank my supervisor, Vincent Sterk, for his guidance and advice. I also thank Raffaella Giacomini and Toru Kitagawa for the opportunity to work closely with them on several projects. I greatly enjoyed our collaborations and I learned a lot from them about doing research. I also thank the academic staff at UCL more broadly for teaching me a heap of stuff over the course of my studies. One of the most important things they (implicitly) taught me is that there is still so much that I know so little about.

Thanks to my fellow MSc/MRes/MPhil/PhD students for making my experience at UCL so great. Special thanks to Hugo, Nathan, Mikkel, Alex, Elena and Cris for many fun nights, quite a few afternoons and even the occasional morning. It was a shame that the fun was truncated by the pandemic, but what can you do?

I thank my parents, Stacy and Paul, for their support over many years. I also apologise to them for disappearing to the other side of the world for five years.

I would be in big trouble if I did not thank my (beautiful) wife, Sandra, for her love and support while undertaking this PhD. I certainly could not have completed it without her. I doubt I would even have started it if it wasn't for her desire to live and work overseas. Thanks for pushing me out of my comfort zone.

The birth of my daughter, Madeline, towards the end of my studies constituted a large negative shock to productivity. However, she is pretty cute, so I don't hold it against her.

Finally, I gratefully acknowledge financial support from the Reserve Bank of Australia, whose views this dissertation does not represent.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A key challenge in modern macroeconomics is understanding the effects of structural shocks on the macroeconomy. This dissertation considers this problem from different perspectives.

In Chapter 2, I explore how monetary policy shocks affect the macroeconomy through the effects on the entry and exit decisions of firms (or ‘firm dynamics’). I document that the firm exit rate increases and the firm entry rate decreases following a contractionary monetary policy shock. Moreover, these margins contribute considerably to the decline in aggregate employment that occurs following the shock. I explore whether these dynamics are important for understanding monetary policy transmission through the lens of a relatively simple heterogeneous-firm New Keynesian model. The model is able to qualitatively match the responses of entry and exit rates following a monetary policy shock. However, accounting for the entry and exit margins has little effect on the predictions of the model relative to those of a representative-agent benchmark. I discuss the forces underlying this approximate equivalence.

In Chapter 3, I develop algorithms to facilitate Bayesian inference in structural vector autoregressions (SVARs) that are set-identified using a mix of sign and zero restrictions. To do this, I show that a system of sign and zero restrictions is equivalent to a system of sign restrictions in a lower-dimensional space. Consequently, algorithms applicable under sign restrictions can be extended to allow for zero restrictions. Specifically, I extend algorithms proposed in Amir-Ahmadi and

Drautzburg (2021) to check whether the identified set is nonempty and to sample from the identified set without rejection sampling. I compare the new algorithms to alternatives by using them to estimate the effects of US monetary policy. The new algorithms are particularly useful when a large number of sign restrictions substantially truncate the identified set given the zero restrictions.

The final three chapters are coauthored with Raffaella Giacomini and Toru Kitagawa. In Chapter 4, we review the literature on robust Bayesian analysis as a tool for global sensitivity analysis and for statistical decision-making under ambiguity. We discuss the methods proposed in the literature, including the different ways of constructing the set of priors that are the key input of the robust Bayesian analysis. We consider both a general set-up for Bayesian statistical decisions and inference and the special case of set-identified structural models. We provide new results that can be used to derive and compute the set of posterior moments for sensitivity analysis and to compute the optimal statistical decision under multiple priors. The paper ends with a self-contained discussion of three different approaches to robust Bayesian inference for set-identified SVARs, including details about numerical implementation and an empirical illustration. This discussion should assist practitioners in estimating the effects of structural shocks when available identifying restrictions are only set-identifying and prior information is at best partially credible.

In Chapter 5, we develop methods for robust Bayesian inference in SVARs where the parameters of interest are set-identified using external instruments, or 'proxy SVARs'. Set-identification in these models typically occurs when there are multiple instruments for multiple structural shocks. Existing Bayesian approaches to inference in proxy SVARs require researchers to specify a single prior over the model's parameters, but, under set-identification, a component of the prior is never revised. We extend robust Bayesian inferential procedures that allow researchers to relax potentially controversial point-identifying restrictions without having to specify an unresiable prior to the case of proxy SVARs. We provide new results on the frequentist validity of the approach in proxy SVARs. We also explore the effect of instrument strength on inference about the identified set. We illustrate our approach by revisiting Mertens and Ravn (2013) and relaxing the assumption that they impose to obtain point identification.

In Chapter 6, we consider SVARs subject to ‘narrative restrictions’, which are inequality restrictions on functions of the structural shocks in specific periods. These restrictions raise novel problems related to identification and inference, and there is currently no frequentist procedure for conducting inference in these models. We propose a solution that is valid from both Bayesian and frequentist perspectives by: 1) formalizing the identification problem under narrative restrictions; 2) correcting a feature of the existing (single-prior) Bayesian approach that can distort inference; 3) proposing a robust (multiple-prior) Bayesian approach that is useful for assessing and eliminating the posterior sensitivity that arises in these models due to the likelihood having flat regions; and 4) showing that the robust Bayesian approach has asymptotic frequentist validity. We illustrate our methods by estimating the effects of US monetary policy under a variety of narrative restrictions.

Chapter 2

Monetary Policy and Firm Dynamics

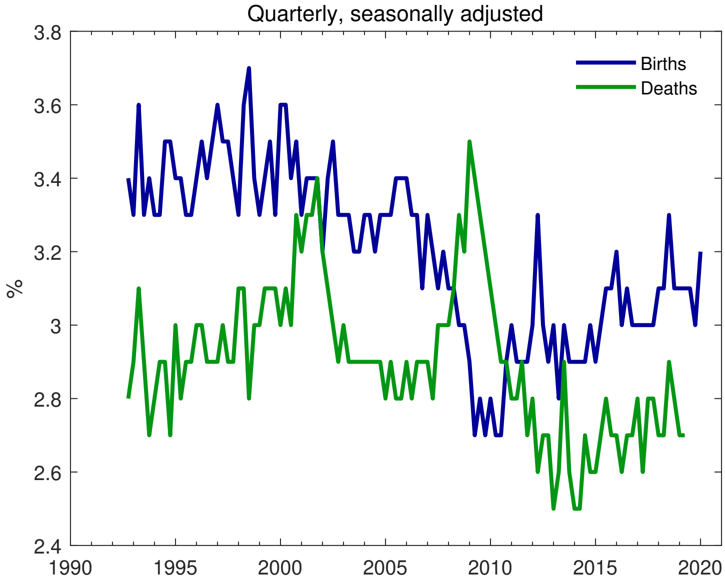
2.1 Introduction

Substantial rates of firm entry and exit are distinct features of the US economy; for example, quarterly entry and exit rates have each averaged around 3 per cent since 1990 and have undergone clear cyclical fluctuations (Figure 2.1).[1] Additionally, these margins contribute appreciably to flows into and out of employment. In this paper, I provide empirical evidence suggesting that the entry and exit behaviour of firms may play an important role in monetary policy transmission. Overall, my results suggest that the primary extensive-margin response of the economy to a contractionary monetary policy shock is via an increase in firm exit. Moreover, I document that changes in job creation and destruction due to changes in entry and exit are sizeable relative to the change in aggregate employment following the shock. I explore the ability of a New Keynesian firm-dynamics model to explain the responses of firm entry and exit and investigate how firm dynamics alter the effects of monetary policy relative to a representative-firm benchmark.

The empirical evidence in this paper is based on a proxy structural vector autoregression (SVAR) in which the monetary policy shock is identified using high-frequency monetary policy surprises (as in Gertler and Karadi (2015)). Estimates from the proxy SVAR suggest that the primary extensive-margin response of the

[1] I refer to 'establishments' and 'firms' interchangeably, although in reality a firm may operate multiple establishments. Data used in the paper are for establishments rather than firms. I use 'firm dynamics' to refer to the endogenous entry and exit of heterogeneous firms and associated changes in the productivity distribution.

Figure 2.1: US Establishment Birth and Death Rates



Notes: Data are from US Bureau of Labor Statistics Business Employment Dynamics

economy following a contractionary monetary policy shock occurs through an increase in firm exit. The response of the entry rate is imprecisely estimated, but the point estimate indicates that the entry rate declines by a smaller magnitude than the change in the exit rate. Back-of-the-envelope calculations suggest that the extensive margin may contribute substantially to the change in aggregate employment following the shock; for example, the increase in job destruction due to higher firm exit in the four years after the shock is equivalent to almost 30 per cent of the change in employment, while the decline in job creation due to lower firm entry is equivalent to almost 20 per cent of the change in employment. The responses of firm entry and exit therefore result in a reduction in employment equivalent to almost half of the total decline in employment over this horizon.

To explore the implications of firm entry and exit for monetary policy transmission, I develop a New Keynesian firm-dynamics model that qualitatively matches the empirical responses of entry and exit to a monetary policy shock. The model incorporates heterogeneous firms in the spirit of Hopenhayn (1992) and Hopenhayn and Rogerson (1993) into an otherwise textbook New Keynesian model (e.g., Galí (2008)). In the model, heterogeneous firms experience idiosyncratic productivity shocks and face fixed entry and operating costs, which induce them to endogenously enter and exit production. These firms sell their output to intermediate goods producers, who set prices subject to a nominal rigidity. This structure for production generates a New Keynesian Phillips Curve linking the behaviour of the heterogeneous firms to aggregate inflation dynamics through the relative price of the heterogeneous firms' output, which is the real marginal cost faced by intermediate goods producers. To focus on the role of entry and exit in the transmission of monetary policy, the model abstracts from firm-level frictions. I calibrate the model to match features of the US economy and explore the model's predictions about the effects of a monetary policy shock.

In the model, a contractionary monetary policy shock causes an increase in the exit rate on impact and a smaller-magnitude decline in the entry rate, which is qualitatively consistent with the empirical evidence. The shock affects entry and exit rates through its effects on the three relative prices that enter the heterogeneous firms' decision problems. Focusing on the response of the exit rate, an increase in the real interest rate means that firms place less value on continuing to operate when deciding whether to pay the fixed operating cost, which makes them more likely to exit. The shock also sets off declines in the relative price of the heterogeneous firms' output (making exit more likely) and the real wage (making exit less likely). The effects of the declines in these two relative prices largely offset one another. Consequently, the overall effect of the shock on the exit rate is predominately driven by the direct effect of the higher real interest rate.

The responses of entry and exit rates to the monetary policy shock induce an endogenous change in aggregate total factor productivity (TFP), which is a channel of monetary policy transmission that is absent in the standard New Keynesian model.[2] The decrease in the entry rate and the increase in the exit rate reduce the total measure of operating firms, which pushes down aggregate TFP (due to decreasing returns to scale). At the same time, the entry and exit rates of less-productive firms are more sensitive to the monetary policy shock than those of more-productive firms, so the changes in entry and exit rates result in an upwards shift in the pro-

[2]Moran and Queralto (2018) develop a New Keynesian model in which TFP is endogenously determined by the decisions of firms to invest in research and development.

ductivity distribution, which pushes up aggregate TFP. The former effect more than offsets the latter, so aggregate TFP falls. Changes in entry and exit rates of similar magnitudes to those estimated in the data result in an extremely persistent, albeit quantitatively small, change in aggregate TFP.

Although changes in entry and exit rates serve to endogenously amplify and propagate the monetary policy shock, impulse responses in the firm-dynamics model are almost identical to those in a comparable representative-firm model. This result is partly due to the quantitatively small responses of entry and exit rates to the shock, but also reflects the small size of firms that are induced to exit (or not enter) due to the shock. Together, these factors mean that the additional change in labour demand due to the extensive margin of adjustment is insubstantial relative to the change in labour demand due to changes in employment along the intensive margin. Furthermore, the role of entry and exit in amplifying the effects of the shock is dampened by changes in prices in general equilibrium. As a result, there is little additional amplification of the shock due to the entry and exit margins. I show that this result is robust to alternative calibrations and assumptions about the structure of the model. Finally, I discuss how the presence of firm-level frictions may increase the roles of entry and exit in the transmission of monetary policy.

Relation to literature. This work relates to an existing literature that uses New Keynesian models to explore the roles of entry and exit in transmitting monetary policy shocks. However, entry and exit in this literature are typically interpretable as the introduction or obsolescence of different varieties of consumption goods produced by monopolistically competitive firms that are otherwise homogeneous (e.g., Lewis (2006), Bilbiie et al. (2007), Bergin and Corsetti (2008) and Bilbiie et al. (2014)). Importantly, these models do not include any meaningful heterogeneity or dynamics at the firm level, which are features of firms' behaviour that have been well-documented using micro data (e.g., Foster et al. (2001)). These models therefore cannot speak to the role of endogenous and persistent changes in the distribution of firms in amplifying and propagating monetary policy shocks. In contrast, firms in my model are heterogeneous with respect to their productivity, as in Hopenhayn (1992) and Hopenhayn and Rogerson (1993). The model therefore allows the potential for distributional dynamics to play a role in monetary policy transmission.

Other papers incorporate heterogeneous firms into New Keynesian models, but do not consider the roles of entry and exit in transmitting monetary policy shocks. In particular, Ottonello and Winberry (2020) develop a heterogeneous-firm New Keynesian model to explore the investment channel of monetary policy. Their model uses a similar structure for production to the model in this paper, in the sense that the source of the nominal rigidity is separated from the heterogeneous firms, but it also includes physical capital and financial frictions. Firms in their model may endogenously default, but the entry process is exogenous and the total measure of operating firms is fixed. Adam and Weber (2019) develop a New Keynesian model in which heterogeneous firms make price-setting decisions subject to a nominal rigidity, but entry and exit in their model is exogenous. By abstracting from firm-level frictions, I provide a useful benchmark against which richer models can be compared. To the best of my knowledge, this is the first paper to document the (near) irrelevance of firm dynamics in an otherwise-standard New Keynesian framework.

More generally, this paper relates to the literature exploring how firm dynamics affect the transmission of macroeconomic shocks. For example, Samaniego (2008), Clementi and Palazzo (2016) and Lee and Mukoyama (2018) consider the effects of aggregate technology shocks in models with heterogeneous firms and endogenous entry and exit, but without nominal rigidities. My finding that firm dynamics are essentially irrelevant for the transmission of monetary policy is similar in spirit to the results in Thomas (2002) and Khan and Thomas (2008), who find that lumpy investment is quantitatively irrelevant in general-equilibrium business-cycle models (without nominal rigidities). My findings also resonate with the 'near-aggregation' result in the heterogeneous-household model of Krusell and Smith (1998).

My empirical results also contribute to a growing body of evidence linking monetary policy to changes in firms' entry and exit behaviour. Bergin and Corsetti (2008), Lewis and Poilly (2012) and Uusküla (2016) document that different measures of firm entry and exit respond to monetary policy shocks using SVARs in which the monetary policy shock is identified by assuming that variables such as output and prices do not respond contemporaneously to the shock. In contrast, I estimate the effects of a monetary policy shock on firm entry and exit using high-frequency monetary policy surprises in a proxy SVAR, as in Gertler and Karadi (2015). The advantage of this approach is that it allows all variables in the system to respond contemporaneously to the monetary policy shock and thus does not require questionable zero restrictions on the contemporaneous relationships among the variables in the VAR. Additionally, I provide new estimates that quantify the importance of firm entry and exit in driving changes in employment following a monetary policy shock.

Outline. The remainder of the paper is structured as follows. Section 2.2 presents new time-series evidence about the responses of firm entry and exit to monetary policy shocks. Section 2.3 develops a New Keynesian firm-dynamics model. Section 2.4 describes the calibration of the model and explores how a monetary policy shock affects the economy in the firm-dynamics model relative to in a comparable representative-firm model. This section also documents the robustness of the model's predictions to changes in the calibration and under alternative assumptions about the model's structure. Section 2.5 concludes.[3]

2.2 Evidence from a Proxy SVAR

In this section, I provide new evidence about the effects of US monetary policy on firm entry and exit using a proxy SVAR (e.g., Mertens and Ravn (2013)). I follow Gertler and Karadi (2015) by using high-frequency surprises in three-month-ahead fed funds futures as a proxy for the monetary policy shock.[4] The motivation behind this proxy is that any change in fed funds futures prices in a narrow window around Federal Open Market Committee announcements should only reflect unexpected news about the path of the federal funds rate rate due to the monetary policy an-

[3]A full set of replication files are available at `https://sites.google.com/view/matthewread/`.

[4]The series of monetary policy surprises is taken from Jarociński and Karadi (2020), which is an updated version of the series constructed by Gürkaynak et al. (2005).

nouncement. The surprises should therefore be correlated with the monetary policy shock and are plausibly exogenous with respect to other structural shocks, which is sufficient to point-identify impulse responses to the monetary policy shock.

The reduced-form VAR specification is a quarterly analogue of the VAR in Gertler and Karadi (2015) that has been augmented with variables measuring firm entry and exit rates. As measures of activity and prices, I include real GDP and the GDP deflator (in logs). The VAR includes the one-year Treasury yield as a measure of the stance of monetary policy instead of the federal funds rate; this is to exploit variation in interest rates due to forward guidance, particularly during the period when the federal funds rate was constrained by the zero lower bound. The VAR also includes the excess bond premium from Gilchrist and Zakrajšek (2012) to control for the systematic response of monetary policy to financial conditions.[5] I add establishment birth and death rates from the US Bureau of Labor Statistics' Business Employment Dynamics (BED) dataset as measures of entry and exit, respectively.[6] The model includes four lags and a constant. The sample begins in the September quarter 1992, which reflects the availability of data on entry and exit, and ends in the December quarter 2016, which reflects the availability of the monetary policy surprises.

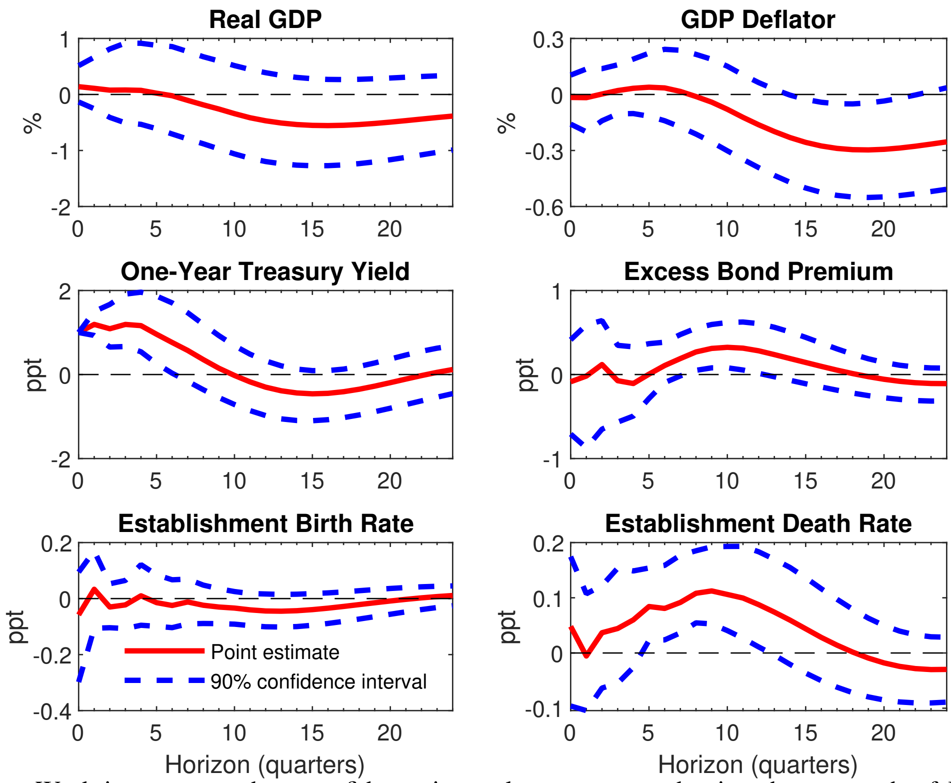
Figure 2.2 plots impulse responses to a monetary policy shock that increases the one-year Treasury yield by 100 basis points on impact. The monetary policy shock results in a persistent increase in the one-year Treasury yield and sluggish falls in output and prices. The peak decline in output is about 0.5 per cent and occurs after four years, although the 90 per cent confidence intervals include zero at all horizons. After around five years, the GDP deflator has declined by about 0.3 per

[5]The excess bond premium is the component of a measure of corporate credit spreads that is orthogonal to predicted default rates and which has been shown to contain predictive information about economic activity. I use the update of this series constructed by Favara *et al.* (2016).

[6]An establishment is an economic unit producing goods or services, typically at one location, and undertaking (mainly) one activity. An establishment birth is an establishment reporting positive employment in the third month of the quarter and zero employment in the third month of the previous four quarters. An establishment death is identified as an establishment reporting zero employment in the third month of four consecutive quarters following a quarter with positive employment (the death is recorded in the first of the four quarters with zero employment). Establishment birth and death rates are measured as a percentage of the average number of establishments in the previous and current quarters. See Sadeghi (2008) for further details.

cent. These responses are consistent with predictions from standard macroeconomic theory. The excess bond premium increases, which is consistent with a tightening in financial conditions.

Figure 2.2: Impulse Responses to Monetary Policy Shock in Proxy SVAR



Notes: Weak-instrument robust confidence intervals are computed using the approach of Montiel Olea et al. (2021).

Following the shock, the quarterly establishment exit rate increases by about 10 basis points over the first two years and the confidence intervals exclude zero at horizons between one and three years. The entry rate declines by a maximum of about 4–5 basis points after around three years, although the confidence intervals include zero at all horizons.[7] The results are qualitatively similar when including the number of entering and exiting establishments (in logs) rather than entry and exit rates. In response to a contractionary monetary policy shock, the number of exiting establishments increases by 1.4 per cent after about two years and the confidence intervals exclude zero at horizons between one and two years after the shock.

[7] These findings are broadly consistent with those in Uusküla (2016), who identifies the shock using a partial causal ordering with different variables in the VAR.

The number of entering establishments decreases by 0.9 per cent after three years, although the confidence intervals include zero at all horizons.

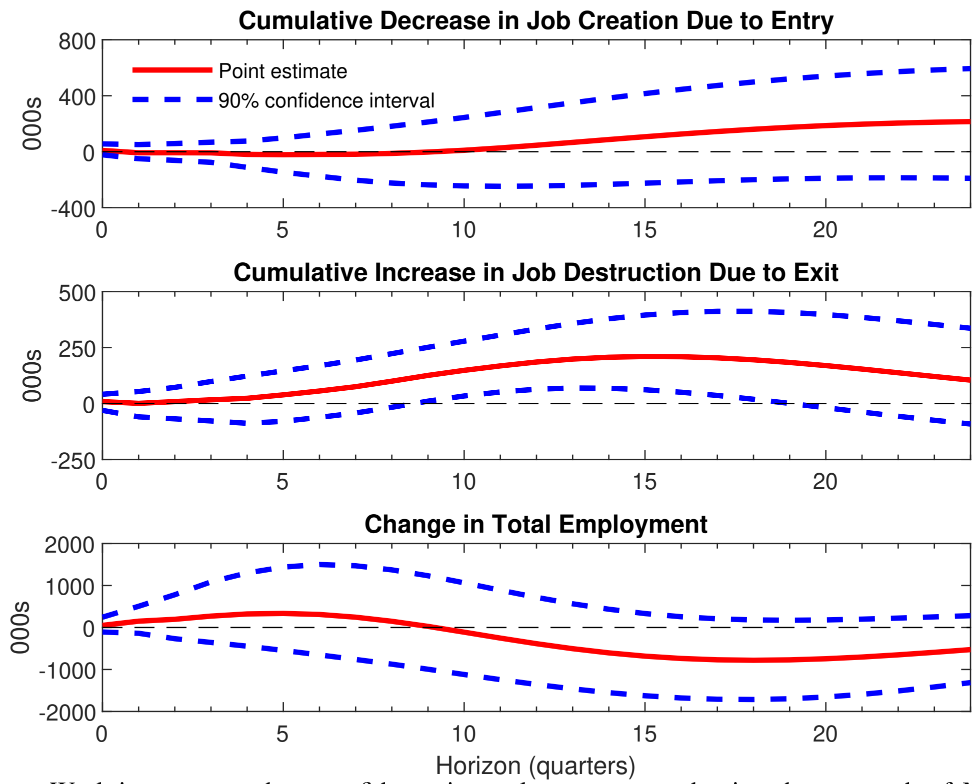
To roughly quantify the contribution of firms' entry and exit behaviour to monetary policy transmission, I replace real GDP with the log of total nonfarm employment in the proxy SVAR that includes entry and exit in levels, and compute the approximate change in employment that is directly due to changes in the entry and exit margins following the shock. I assume that variables are initially at their December 2016 levels, and compute the contribution of the entry (exit) margin to the change in employment by cumulating the change in the number of entrants (exits) and multiplying by the average size of an entrant (exit) in the Business Dynamics Statistics (BDS).[8] This exercise assumes that firms that do not enter (or that exit) due to the shock are on average the same size as entrants (or exiting firms) unconditionally.

Figure 2.3 summarises the results from this exercise. Four years after the shock, employment has declined by about 0.5 per cent, which is equivalent to about 740,000 workers. The increase in firm exit due to the shock means that an additional 210,000 jobs are destroyed in the four years following the shock, which is equivalent to almost 30 per cent of the total decline in employment. At the same time, the decrease in firm entry due to the shock means that around 125,000 fewer jobs are created in the four years following the shock, which is equivalent to a bit less than 20 per cent of the total decline in employment. Note, however, that the confidence intervals for the cumulative change in employment attributable to the entry margin include zero at all horizons. The point estimates suggest that the combined response along the extensive margin directly results in a reduction in employment that is equivalent to almost half of the total decline in employment four years after the shock.

Overall, the results from the proxy SVAR indicate that a contractionary mon-

[8]The BDS are publicly released statistics aggregated from the US Census Bureau's Longitudinal Business Database. In this exercise, I assume that the initial level of employment is 145 million. The initial numbers of entering and exiting firms per quarter are, respectively, 239,000 and 217,000. Based on the 2016 release of the BDS, the average number of workers employed by entering firms is 8.2 and by exiting firms is 7.7.

Figure 2.3: Employment Effects of Monetary Policy Shock in Proxy SVAR



Notes: Weak-instrument robust confidence intervals are computed using the approach of Montiel Olea et al. (2021).

etary policy shock results in an increase in firm exit and a decrease in firm entry, although impulse responses of the measures of entry are not significantly different from zero at conventional levels of significance. The exit margin appears to be more responsive to a monetary policy shock than the entry margin, so the primary extensive-margin response of the economy to a monetary policy shock appears to be via firm exit rather than firm entry. Moreover, the estimates suggest that these margins may contribute significantly to the total change in employment following a monetary policy shock.

2.3 A New Keynesian Firm-Dynamics Model

In this section, I develop a New Keynesian model with heterogeneous firms that endogenously enter and exit production. The model embeds the firm-dynamics model of Hopenhayn (1992) and Hopenhayn and Rogerson (1993) into a textbook New Keynesian model (e.g., Galí (2008)).

Time in the model is discrete (indexed by $t = 0, 1, \ldots$) and all agents have rational expectations. For tractability, I follow Ottonello and Winberry (2020) by segmenting production into different stages to separate the source of nominal rigidity from the heterogeneous firms' production decisions. 'Production' firms competitively produce an undifferentiated good using labour as the sole input in a production function with decreasing returns to scale. These firms face idiosyncratic TFP shocks and fixed entry and operating costs, which induce them to enter and exit production. Intermediate goods producers purchase the undifferentiated good from production firms and use it to produce a differentiated good, which they sell in a monopolistically competitive market subject to a nominal rigidity. Final goods producers purchase the differentiated goods and bundle them into a final good. Households enjoy utility from leisure and consumption of the final good, and can save in a nominal bond that is in zero net supply. The central bank sets the interest rate on the nominal bond according to a Taylor rule. The government collects fixed costs from the production firms and distributes these to the household as a lump-sum transfer.

In the following subsections, I detail the decision problems faced by each type of agent and develop the equilibrium conditions.

2.3.1 Households

A representative household enjoys utility from consumption (C_t) and suffers disutility from supplying labour (N_t) at real wage rate w_t. The household has access to a nominal bond (B_t) paying gross nominal interest rate R_t. The household earns real dividends (D_t) from owning firms and receives real lump-sum transfers (T_t) from the government. The household's utility maximisation problem is

$$\max_{\{C_t, N_t, B_t\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma} - 1}{1-\sigma} - \kappa_0 \frac{N_t^{1+\kappa_1}}{1+\kappa_1} \right) \tag{2.1}$$

subject to the budget constraint

$$C_t + \frac{B_t}{P_t} = R_{t-1} \frac{B_{t-1}}{P_t} + w_t N_t + D_t + T_t, \tag{2.2}$$

where P_t is the price of the final good, β is the subjective discount factor, κ_0 scales the disutility from supplying labour and κ_1 is the inverse of the Frisch elasticity of labour supply.

The first-order conditions for the problem imply the labour supply condition

$$w_t = \kappa_0 C_t^{\sigma} N_t^{\kappa_1} \tag{2.3}$$

and the Euler equation

$$1 = \mathbb{E}_t \left(\Lambda_{t,t+1} \frac{R_t}{\Pi_{t+1}} \right), \tag{2.4}$$

where $\Pi_{t+1} = P_{t+1}/P_t$ is the gross rate of inflation between periods t and $t+1$, and $\Lambda_{t,t+k} = \beta^k (C_{t+k}/C_t)^{-\sigma}$ is the representative household's k-period stochastic discount factor.

2.3.2 Central bank and government

The central bank sets the nominal interest rate based on an inflation-targeting rule:

$$\frac{R_t}{R} = \left(\frac{\Pi_t}{\Pi} \right)^{\phi} \exp(\varepsilon_t^m), \tag{2.5}$$

where R is the nominal interest rate in the stationary equilibrium and Π is the inflation target. ε_t^m is a monetary policy shock following the AR(1) process $\varepsilon_t^m = \rho_m \varepsilon_{t-1}^m + \eta_t$, where η_t is white noise.[9]

The only role of the government is to collect entry and operating costs paid by production firms and remit these to the household as a lump-sum transfer, which will mean that the fixed costs paid by production firms do not appear in any market clearing conditions. This facilitates comparison against the representative-firm model, in which there are no fixed costs.

[9] As in Ottonello and Winberry (2020), including persistence in the monetary policy shock is a simple way to induce persistence in the responses of variables to the shock. An alternative would be to include interest-rate smoothing in the Taylor rule.

2.3.3 Production firms

The structure for production of the undifferentiated good is similar to that in Hopenhayn and Rogerson (1993). There is a measure of heterogeneous production firms (indexed by j) producing an undifferentiated good according to the production function

$$y_{jt} = z_{jt} n_{jt}^{\nu}, \tag{2.6}$$

where $z_{jt} \in \mathbb{R}_+$ is idiosyncratic productivity, n_{jt} is the firm's labour input, and $0 < \nu < 1$, so there are decreasing returns to scale. z_{jt} follows an AR(1) process in logs:

$$\ln z_{jt} = a_z(1-\rho_z) + \rho_z \ln z_{j,t-1} + \varepsilon_{jt}^z, \quad \varepsilon_{jt}^z \overset{iid}{\sim} N(0, \sigma_z^2). \tag{2.7}$$

The cumulative distribution function (CDF) of z_{jt} conditional on $z_{j,t-1}$ is $F(.|z_{j,t-1})$.

As in Clementi and Palazzo (2016), production firms face a stochastic fixed operating cost, c_{jt}, denominated in units of the final good. The fixed cost is independently and identically distributed across firms and over time according to a lognormal distribution: $c_{jt} \overset{iid}{\sim} LN(\mu_c, \sigma_c)$ with CDF $G_c(.)$. At the beginning of each period, firms learn the current draw of z_{jt}, choose n_{jt} and produce output. Firms then draw the fixed operating cost and must choose whether to pay this cost in the current period to continue operating in the next period or to shut down and earn zero profits thereafter.

There is also a fixed mass M of potential entrants (indexed by m). At the beginning of each period, potential entrants draw z_{mt} from a distribution with CDF $Q(.)$. Potential entrants who decide to begin operating incur a one-time fixed cost, e_{mt}, denominated in units of the final good. Unlike in Hopenhayn and Rogerson (1993), this cost is stochastic; in particular, $e_{mt} \overset{iid}{\sim} \times LN(\mu_e, \sigma_e)$ with CDF $G_e(.)$. Potential entrants draw the entry cost after drawing z_{mt} but before deciding whether to enter. Entrants then operate in exactly the same way as an existing firm. Entry and operating costs are financed using equity.

The individual state variable for an operating firm is z_{jt}. The aggregate state variables are the distribution of incumbent firms over idiosyncratic productivity (described below) and the aggregate shock ε_t^m. Firms care about the aggregate state of the economy only to the extent that it affects prices, which they take as given. I summarise the aggregate state by including time, t, as an index in the firm's value function. For ease of notation, I henceforth drop firm and time subscripts from firm-specific variables and denote next-period values with a prime.

Production firms operate in a competitive output market and hire labour in a competitive spot market at real wage rate w_t. They discount the future using the representative household's stochastic discount factor. Let $V_t(z)$ be the expected discounted value of future real profits (in units of the final good) for an incumbent firm entering period t with idiosyncratic productivity z and who behaves optimally. Consider the decision problem of a firm that has just drawn fixed operating cost c and must decide whether to pay this fixed cost and continue operating in the next period or to exit. The firm will only continue to operate if $\mathbb{E}_t\left(\Lambda_{t,t+1}V_{t+1}(z')\right) \geq c$. The left-hand side of this inequality implicitly depends on z (because z follows an AR(1) process) and is constant in c, while the right-hand side is increasing in c. There thus exists a threshold value of c at each value of z, $c_t^*(z) = \mathbb{E}_t\left(\Lambda_{t,t+1}V_{t+1}(z')\right)$, below which the firm will continue to operate and above which it will exit.[10]

The Bellman equation for the firm's problem is

$$V_t(z) = \max_{n \geq 0} p_t z n^{\nu} - w_t n + \int \max\left\{\mathbb{E}_t\left(\Lambda_{t,t+1}V_{t+1}(z')\right) - c, 0\right\} dG_c(c), \tag{2.8}$$

where p_t is the relative price of the undifferentiated good in units of the final good. The first-order condition for the problem implies the policy function for labour demand:

$$n_t(z) = \left(\frac{w_t}{\nu p_t z}\right)^{\frac{1}{\nu-1}}. \tag{2.9}$$

[10]If the fixed operating cost were not stochastic, as in Hopenhayn and Rogerson (1993), there would exist a threshold value of idiosyncratic productivity below which all firms would exit and above which all firms would choose to continue operating, which is inconsistent with observed patterns of firm exit.

Using this policy function and the threshold for exit, the Bellman equation becomes

$$V_t(z) = p_t z n_t(z)^{\nu} - w_t n_t(z) + \left[\mathbb{E}_t\left(\Lambda_{t,t+1} V_{t+1}(z')\right) - \mathbb{E}_c\left(c | c \leq c_t^*(z)\right)\right] G_c\left(c_t^*(z)\right). \tag{2.10}$$

The term in square brackets is the expected discounted value of operating in time $t+1$ net of the expected fixed operating cost conditional on this cost being less than the expected discounted value of operating.[11] This is multiplied by the probability that the fixed operating cost is less than the exit threshold. With the complementary probability, the fixed operating cost is greater than the exit threshold, in which case the firm will exit and will have zero continuation value.

Similarly, a potential entrant that has drawn idiosyncratic productivity z and fixed entry cost e will only begin operating if $V_t(z) \geq e$. There exists a threshold value of e at each value of z, $e_t^*(z) = V_t(z)$, below which a potential entrant will choose to enter and above which it will not.

The measure of operating firms across idiosyncratic productivity in period t, $\mu_t(z)$, consists of firms that were operating in period $t-1$ that chose not to exit and new entrants in period t. For all Borel subsets $B \in \mathbb{R}_+$ of the idiosyncratic state space, this measure evolves according to

$$\mu_{t+1}(B) = \int \int_{z' \in B} G_c\left(c_t^*(z)\right) dF(z'|z) d\mu_t(z) + M \int_{z' \in B} G_e\left(e_{t+1}^*(z')\right) dQ(z'). \tag{2.11}$$

This is the relevant measure for computing aggregate output, employment, etc.[12] Aggregate output is $Y_t = \int z n_t(z)^{\nu} d\mu_t(z)$, aggregate employment is $N_t = \int n_t(z) d\mu_t(z)$ and aggregate real profit earned by the production firms is

[11] $\mathbb{E}_c\left(c | c \leq c_t^*(z)\right)$ is the expected value of a truncated lognormal random variable. If $x \sim LN(\mu, \sigma)$ and $\Phi(.)$ is the standard normal CDF, then

$$\mathbb{E}_x(x | x \leq k) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \frac{\Phi\left(\frac{\ln k - \mu - \sigma^2}{\sigma}\right)}{\Phi\left(\frac{\ln k - \mu}{\sigma}\right)}.$$

[12] The measure of operating firms over idiosyncratic productivity is not part of the aggregate state, since it is not predetermined; it depends on the entry decisions of firms, which in turn depend on prices in the current period. The aggregate state includes the measure of *incumbent* firms, which is equal to the measure of operating firms less new entrants.

$\Omega_t = p_t Y_t - w_t N_t - T_t$, where T_t are aggregate fixed operating and entry costs:

$$T_t = \int G_c\left(c_t^*(z)\right) \mathbb{E}_c\left(c | c \leq c_t^*(z)\right) d\mu_t(z) + M \int G_e\left(e_t^*(z)\right) \mathbb{E}_e\left(e | e \leq e_t^*(z)\right) dQ(z). \tag{2.12}$$

2.3.4 Intermediate and final goods producers

Since this part of the model is fairly standard, I describe it at a high level and relegate detail to Appendix A.1. There is a unit mass of intermediate goods producers, which purchase the undifferentiated good from the production firms and costlessly differentiate it. Intermediate goods producers sell their products to a representative final good producer in a monopolistically competitive market, taking the demand schedule of the final good producer as given. Intermediate goods producers face a quadratic price-adjustment cost – paid in units of the final good – as in Rotemberg (1982). The final good producer bundles intermediate goods into a final good according to a CES production function, which generates a downward-sloping demand schedule for the intermediate goods.

I consider symmetric equilibria where all intermediate goods firms face the same initial price, which implies that they will optimally choose the same price in each period. Intermediate firms' optimal price-setting behaviour is characterised by a (nonlinear) Phillips Curve:

$$(1-\gamma) + \gamma p_t - \xi(\Pi_t - 1)\Pi_t = -\xi \mathbb{E}_t \left[\Lambda_{t,t+1}(\Pi_{t+1} - 1)\Pi_{t+1} \frac{Y_{t+1}}{Y_t}\right]. \tag{2.13}$$

Real profit earned by the intermediate goods firms in a symmetric equilibrium is

$$\Upsilon_t = \left(1 - p_t - \frac{\xi}{2}(\Pi_t - 1)^2\right) Y_t. \tag{2.14}$$

Aggregate dividends are then $D_t = \Omega_t + \Upsilon_t$.

2.3.5 Equilibrium

Given the law of motion for the exogenous state ε_t^m, the competitive equilibrium is a joint law of motion for production firms' policy functions and value function

$\{n_t(z), c_t^*(z), e_t^*(z), V_t(z)\}_{z\in\mathbb{R}_+}$, the measure of operating firms over idiosyncratic productivity, $\{\mu_t(z)\}_{z\in\mathbb{R}_+}$, aggregate quantities $\{C_t, N_t, B_t, Y_t, D_t, T_t\}$ and (relative) prices $\{p_t, w_t, \Pi_t\}$ such that, for all t: the value function solves the production firms' Bellman equation for all $z \in \mathbb{R}_+$ with associated policy functions $n_t(z)$, $c_t^*(z)$ and $e_t^*(z)$; the Euler equation and labour supply condition are satisfied; the Taylor Rule and Phillips Curve are satisfied; $\mu_t(z)$ evolves according to the transition function in Equation 2.11; the profits of production firms and intermediate goods producers are paid out as dividends; fixed costs are remitted to the household according to Equation 2.12; and goods, labour and bond markets clear.[13]

2.3.6 Stationary equilibrium

In a stationary equilibrium without aggregate risk, all aggregate quantities and relative prices are constant. The measure of firms across idiosyncratic productivity is also constant, although the position of individual firms will move around within the distribution, and firms will enter and exit.

I consider a stationary equilibrium with zero net inflation, so $\Pi = 1$. The Euler equation implies that $R = 1/\beta$ and intermediate firms' price-setting condition implies that $p = (\gamma - 1)/\gamma$. Given p and an arbitrary value of w, solving the production firms' problem yields the value function and the policy functions for labour demand, entry and exit. The stationary measure of firms is then a fixed point of

$$\mu(B) = \int\int_{z'\in B} G_c\left(c^*(z)\right) dF(z'|z) d\mu(z) + M \int_{z'\in B} G_e\left(e^*(z')\right) dQ(z'), \qquad (2.15)$$

which can be used to compute aggregate output $Y = \int z n(z)^\nu d\mu(z)$ and labour demand $N = \int n(z) d\mu(z)$. Given the value of w and imposing labour market clearing then implies a unique value of consumption C from the labour supply condition. One could then check whether the final good market clears (i.e., $Y = C$). Since the choice of w was arbitrary, clearing of the final good market pins down the real wage

[13] The labour market clearing condition is $N_t = \int n_t(z) d\mu_t(z)$, the bond market clearing condition is $B_t = 0$ (since bonds are in zero net supply) and the (final) goods market clearing condition is $Y_t = C_t + \frac{\xi}{2}(\Pi_t - 1)^2 Y_t$. Clearing of the markets for production and intermediate goods is implicit in the Phillips Curve.

in the stationary equilibrium.[14]

2.4 Quantitative Results

The goal of this section is to explore the ability of the model developed in Section 2.3 to explain the responses of firm entry and exit following a monetary policy shock, and to examine how allowing for firm dynamics alters the effects of the shock relative to a representative-firm benchmark. To this end, I first calibrate the model so that features of its stationary equilibrium broadly match US data. I then solve for the model's equilibrium dynamics in the vicinity of the stationary equilibrium using first-order perturbation.

2.4.1 Calibration

The model period is one quarter. The target for the annual real interest rate is 4 per cent, so $\beta = 1.04^{-1/4}$. $\sigma = 1$, so there is log utility in consumption. $\kappa_1 = 1$, so there is a unit Frisch elasticity. $\gamma = 6$, so intermediate goods producers set a markup of 20 per cent in the stationary equilibrium. $\phi = 1.5$, so the central bank moves the nominal interest rate more than one-for-one in response to deviations of inflation from target. These choices are within the range of values usually considered in the literature. ν, which controls returns to scale, is 0.9, which is within the range of estimates in Basu and Fernald (1997) and Lee (2005). ξ is set to 50 to deliver a Phillips Curve slope of 0.1, as in Ottonello and Winberry (2020).

In a stationary equilibrium, the log of firm-level employment follows the AR(1) process

$$\ln n_{jt} = \frac{1-\rho_z}{\nu - 1} \ln \frac{\nu}{pa_z} + \rho_z \ln n_{j,t-1} + \left(\frac{1}{\nu - 1}\right) \varepsilon^z_{jt}. \tag{2.16}$$

Using annual data from the US Census Bureau's Longitudinal Business Database, Pugsley et al. (2021) estimate the parameters of different processes for firm-level employment – including an AR(1) process – by matching the observed cross-sectional autocovariance structure of employment. Based on a balanced panel of

[14] Under the calibration of the model described below, I have verified numerically that there is a unique stationary equilibrium; aggregate supply of the final good is decreasing in w and, given that the labour market clears, aggregate demand for the final good is increasing in w.

firms, they estimate an AR(1) persistence parameter of $\rho_n = 0.9771$ and an innovation standard deviation of $\sigma_n = 0.2676$.[15] The implied persistence parameter in the quarterly process for log productivity is $\rho_z = \rho_n^{1/4}$ and the implied innovation standard deviation is $\sigma_z = (\nu - 1)\sqrt{\sigma_n^2 / \sum_{j=0}^{3} \rho_z^{2j}}$.

I discretise the process for idiosyncratic productivity using the method of Rouwenhorst (1995) (as described in Kopecky and Suen (2010)) with 50 evenly spaced points for $\ln z$ and solve the Bellman equation for production firms on this grid using value function iteration. The distribution from which potential entrants draw idiosyncratic productivity, $Q(z)$, is the same as the unconditional distribution faced by existing firms. The distribution of entry costs coincides with the distribution of fixed operating costs, so $\mu_e = \mu_c$, $\sigma_e = \sigma_c$ and $G_e(.) = G_c(.)$. The stationary measure of firms is obtained by iterating on its transition function to convergence.

I normalise the real wage w to unity and find M such that the final good market clears.[16] I jointly calibrate the remaining parameters $(\mu_c, \sigma_c, a_z, \kappa_0)$ to target the annual exit rate (8.6 per cent), the average size of an incumbent firm (19.2 employees), the average size of exiting firms (7.7 employees), and the employment-to-population ratio (0.6). The first three targets are taken from the 2016 BDS, while the last is roughly equal to the 2016 employment-to-population ratio based on the Bureau of Labor Statistics' Current Population Survey.[17] Table 2.1 lists the parameter values under this calibration.

The targeted moments in the model are equal to the moments in the data. The model also appears to perform reasonably well in matching non-targeted features of the distribution of firms, particularly given the simplicity of the model and the small set of calibration targets. As in the data, the model predicts that there are more small firms than large firms, although large firms account for a smaller share of total employment in the model than in the data (Figure 2.4, left panels). The

[15] See column (5) in Table B.2 in their online appendix.

[16] The measure of firms is homogeneous of degree one in M, and thus so is aggregate employment. Given targets for the average size of incumbent firms and the employment-to-population ratio, M essentially adjusts so that the measure of firms reconciles these two moments (at $w = 1$).

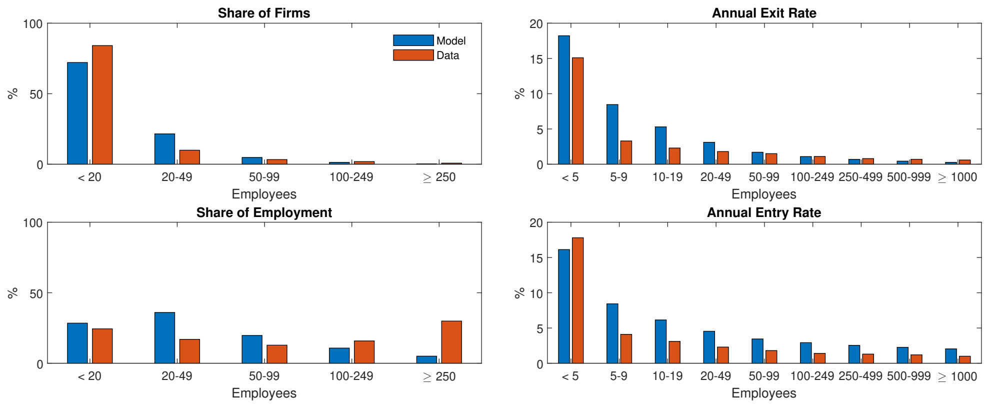
[17] I estimate the average size of an incumbent firm in the BDS by subtracting jobs created due to establishment births from total employment and dividing by the total number of establishments less the number of establishment births.

Table 2.1: Calibrated Parameters

Parameter	Value	Parameter	Value
β	0.99	κ_0	2.083
σ	1	κ_1	1
ν	0.9	ϕ	1.5
γ	6	ξ	50
μ_c	–6.216	σ_c	4.537
σ_z	0.013	ρ_z	0.994
M	7.483×10^{-4}	a_z	0.439

profiles of entry and exit rates in the model and data are similar in the sense that entry and exit rates decline with firm size (Figure 2.4, right panels). The model also closely matches the age distribution of firms and the share of employment by age in the data.

Figure 2.4: Distribution of Firms by Size



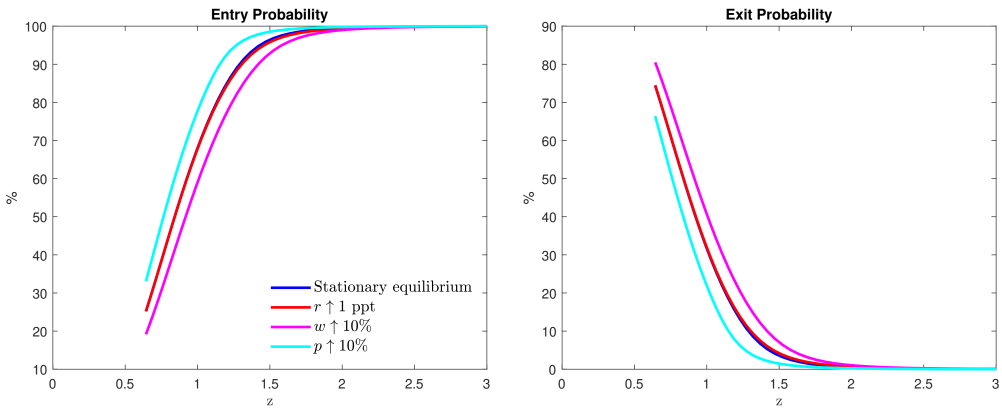
Notes: Data are from 2016 BDS.

To illustrate how the firms' entry and exit decisions depend on prices, Figure 2.5 plots entry and exit probabilities in the stationary equilibrium as a function of idiosyncratic productivity, along with these probabilities when one price at a time is varied.[18] All else equal, a higher real interest rate (r) increases the exit probability across the productivity distribution, because firms place less weight on future

[18]The entry probability (before observing the draw of the entry cost) is the probability that a potential entrant with productivity draw z chooses to enter (i.e., $G_e(e_t^*(z))$); this probability is distinct from the entry *rate*, which is the mass of actual entrants at each level of productivity divided by the measure of firms operating at that level of productivity. The exit probability is the probability (before observing the draw of the fixed operating cost) that an operating firm with productivity draw z chooses to exit (i.e., $1 - G_c(c_t^*(z))$); in the stationary equilibrium, this probability coincides with the exit *rate*.

profits when comparing their draw of the fixed operating cost against the expected value of continuing to operate. A higher real wage (w) directly decreases firm profits and thus decreases the value of continuing to operate, and so exit probabilities are higher. Conversely, a higher relative price of production firms' output (p) directly increases firm profits and yields lower exit probabilities. Similar reasoning applies to the profile of entry probabilities. Entry and exit probabilities tend to be more sensitive to changes in prices at lower levels of productivity than at higher levels, which reflects differences in the effect of price changes on the value of operating as well as the shape of the distribution of fixed costs.

Figure 2.5: Entry and Exit Probabilities by Idiosyncratic Productivity



2.4.2 Dynamic effects of a monetary policy shock

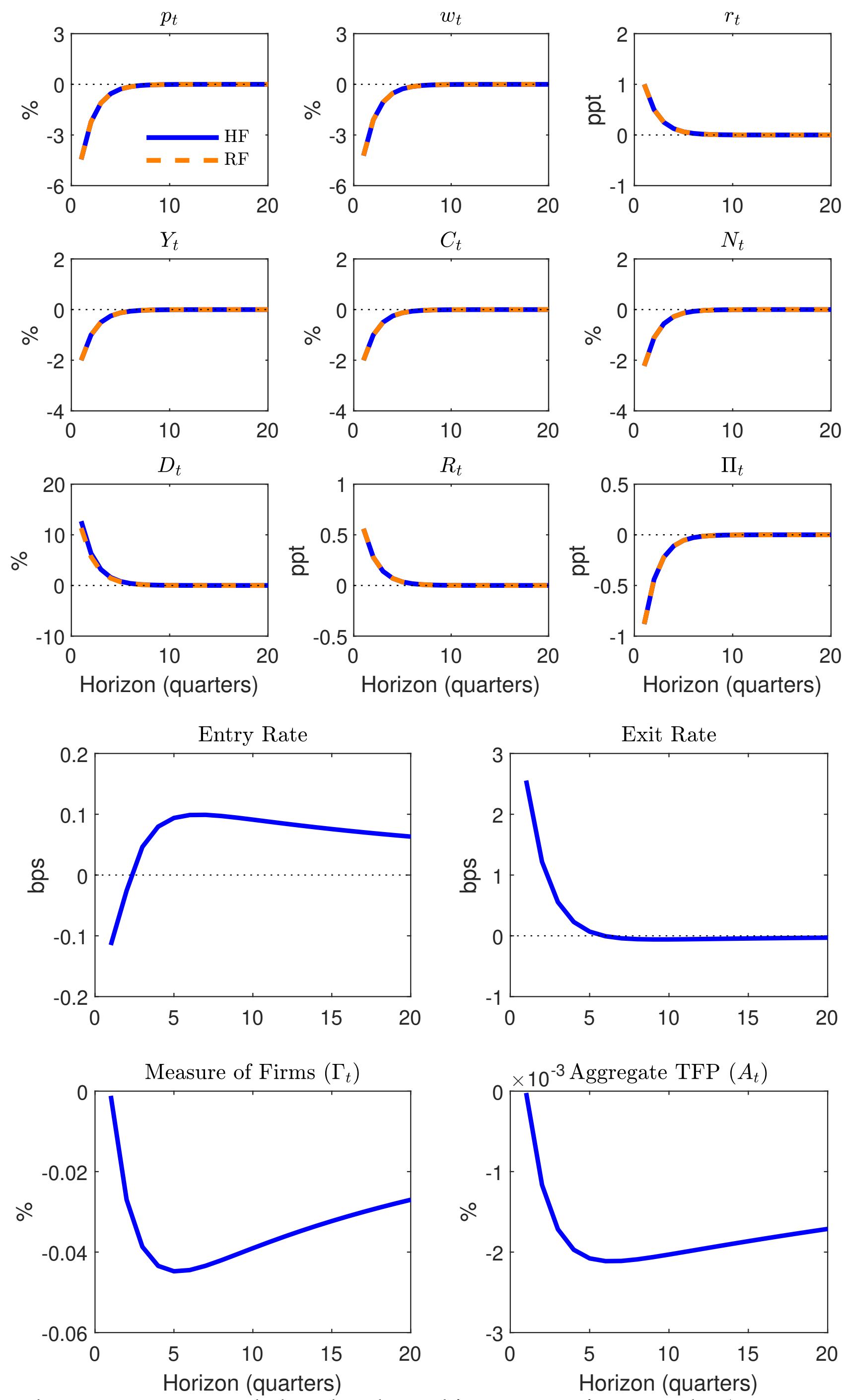
I trace out the effects of a monetary policy shock in the calibrated model by solving for the model's equilibrium dynamics in the vicinity of the stationary equilibrium using a first-order perturbation solution method (see Appendix A.1 for details). I assume that the parameter in the AR(1) process for the monetary policy shock (ρ_m) is 0.5, which is the same value used in Ottonello and Winberry (2020). I qualitatively compare the responses of entry and exit rates in the heterogeneous-firm (HF) model against those obtained using the proxy SVAR in Section 2.2, and explore the role of firm dynamics in transmitting the shock by comparing impulse responses against those from a comparable representative-firm (RF) model. The RF model replaces the mass of heterogeneous production firms with a representative firm that chooses labour input to maximise profits subject to a production function with decreasing returns to scale. The firm operates in competitive output and labour input markets and does not pay a fixed operating cost. In Appendix A.2, I show that this RF model is identical to the standard three-equation New Keynesian model, except that decreasing returns to scale in production induce a steeper Phillips Curve than under a linear production technology. The parameter values in the calibrated RF model are the same as those in the HF model.

Summary of responses. Figure 2.6 plots impulse responses to a positive monetary policy shock scaled so that the *ex ante* real interest rate increases by one percentage point on impact. Consumption, output and employment decline in response to the shock, as does the relative price of production firms' output and the real wage. The exit rate increases by about 2.5 basis points on impact before declining towards its stationary-equilibrium value.[19] The entry rate initially declines by about 0.1 basis points, but then increases to be higher than in the stationary equilibrium, which reflects a decline in the denominator due to the increase in the exit rate. The increase in the exit rate, the decrease in the entry rate and the larger response of the exit rate relative to the entry rate are qualitatively consistent with the results from the proxy SVAR presented in Section 2.2, although the responses are clearly quantitatively small. The changes in firm entry and exit induce a small but persistent decrease in the total measure of operating firms (Γ_t), which is still around 0.03 per cent lower after five years.

Notably, the responses of macroeconomic variables to the shock are essentially indistinguishable across the HF and RF models; for example, output declines by 2.0023 per cent in the HF model compared with 2 per cent in the RF model. Allowing for the entry and exit of heterogeneous firms in an otherwise textbook New Keynesian model therefore has no appreciable effect on the responses of most

[19]The definitions of entry and exit rates in this exercise are consistent with measurement in the BED. The exit rate in period t is equal to the measure of firms who choose to exit at the end of period t divided by the measure of operating firms averaged over periods t and $t+1$: $\int(1-G_c(c_t^*(z)))d\mu_t(z)/X_{t+1}$, where $X_{t+1} = 0.5(\Gamma_{t+1}+\Gamma_t)$ and $\Gamma_t = \int d\mu_t(z)$ is the total measure of operating firms. The entry rate in period t is the measure of potential entrants who choose to enter in period t divided by the measure of operating firms averaged over periods t and $t-1$: $M\int G_e(e_t^*(z))dQ(z)/X_t$.

Figure 2.6: Impulse Responses to Monetary Policy Shock in HF and RF Models



Notes: Impulse responses are scaled so that the real interest rate increases by 1 percentage point on impact.

variables to a monetary policy shock.[20] I discuss this result in more detail below.

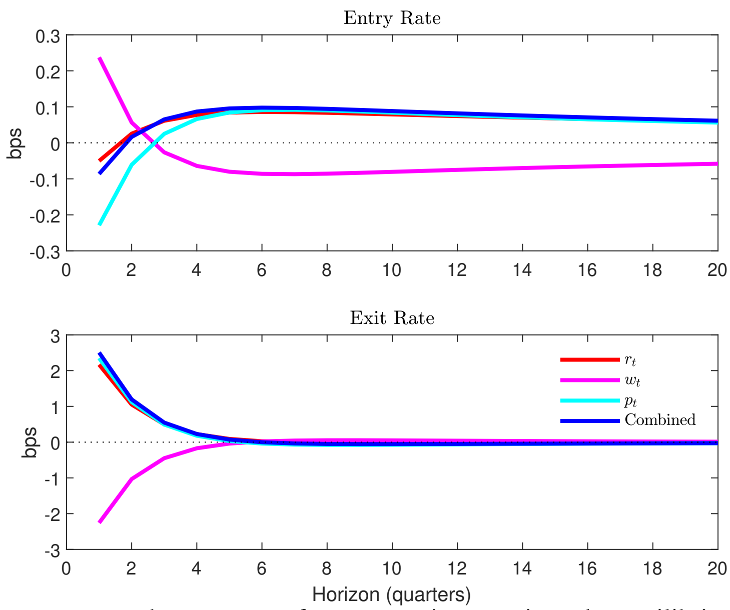
What drives the responses of entry and exit rates? The entry probability at a particular level of idiosyncratic productivity is $G_c(e_t^*(z))$ and the exit probability is $1 - G_c(c_t^*(z))$, where $e_t^*(z) = V_t(z)$ and $c_t^*(z) = \mathbb{E}_t(\Lambda_{t,t+1}V_{t+1}(z))$. The responses of these probabilities to the monetary policy shock will therefore depend on the shape of the fixed-cost distribution, the response of (expected) firm value to changes in the entire sequence of future prices, and the equilibrium changes in prices themselves. The responses of entry and exit rates then depend on the responses of entry and exit probabilities across the distribution of idiosyncratic productivity for potential entrants and operating firms, respectively.

Focusing on the response of the exit rate, the increase in the real interest rate is a decrease in the stochastic discount factor used by production firms to discount future profits. When deciding whether to pay the fixed operating cost to continue operating in the future, the production firms place less value on continuing to operate. Consequently, a larger proportion of firms draw a fixed operating cost that exceeds their exit threshold and, all else equal, the exit rate increases. An increase in the real interest rate also decreases household demand for the final good, which – due to sticky prices – decreases intermediate goods firms' demand for the undifferentiated good. This drives down the relative price of production firms' output (p_t) and thus reduces the value to production firms of operating, which would tend to increase the exit rate. However, because production firms operate in competitive markets, any reduction in the relative price of their output must be met with a decline in real marginal costs. Given decreasing returns to scale in production, this is achieved through production firms reducing employment, which pushes down the real wage (w_t). All else equal, this decline in the real wage would tend to result in a decrease in the exit rate. Similarly, changes in w_t due to changes in labour supply will be transmitted through to changes in p_t. This dynamic induces a strong positive co-movement between p_t and w_t.

[20] An exception is the real dividend, whose increase on impact is about 1.4 percentage points larger in the HF model; this is partly due to a decline in fixed operating costs paid by production firms.

Figure 2.7 plots the paths of the entry and exit rates obtained by feeding the equilibrium response of each price into the firms' problem while holding other prices constant and assuming that firms have perfect foresight about the path of prices. It is evident that the movements in w_t and p_t have effects on entry and exit rates that largely offset one another. Consequently, the responses of entry and exit rates are due largely to the direct effect of the change in the real interest rate.

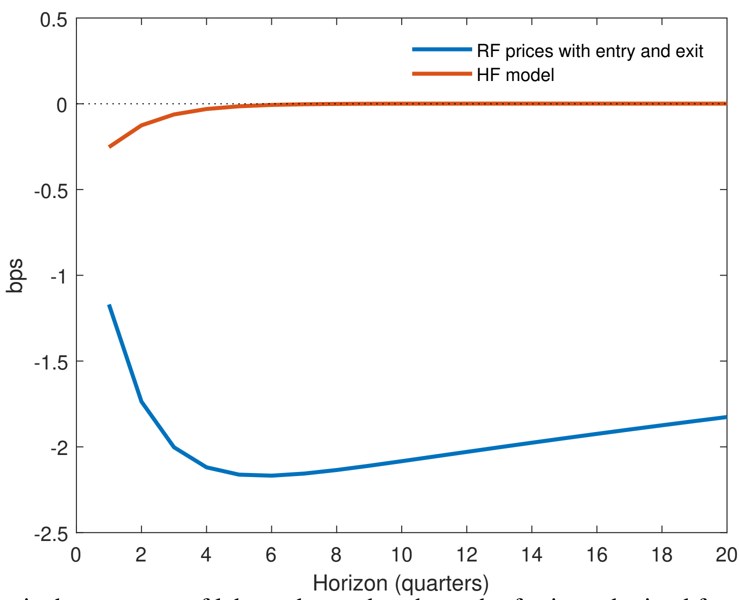
Figure 2.7: Contributions of Prices to Responses of Entry and Exit Rates



Notes: Each line represents the response of entry or exit rates given the equilibrium path of a particular price following the shock, holding other prices constant and assuming that firms have perfect foresight about the path of prices; blue line is the overall response.

Amplification and propagation of monetary policy shocks. In the RF model, a reduction in p_t results in the firm reducing its scale of operation, which reduces labour demand and thus puts downward pressure on the real wage. This intensive-margin adjustment also occurs in the HF model, with the degree of adjustment for each firm depending on the level of idiosyncratic productivity, but there is an additional extensive-margin adjustment; a persistent reduction in p_t reduces the value of continuing, which reduces the threshold value of the fixed operating cost above which firms continue operating, so a greater proportion of firms exit at each level of productivity. Similarly, a reduction in p_t results in a lower proportion of potential entrants entering at each level of productivity. These extensive-margin adjustments result in an additional decline in labour demand. To illustrate, I feed in the paths of prices from the RF model into the decision problem of production firms in the HF model (assuming perfect foresight about prices). This generates a fall in employment that is larger than the fall in employment in the RF model (Figure 2.8, blue line). In the HF model, this additional margin of adjustment puts further downward pressure on the real wage. As discussed above, a reduction in the real wage is a reduction in real marginal cost for production firms, which must be met with a reduction in p_t due to competition in the market for the production good. The entry and exit margins therefore amplify the effects of the monetary policy shock on w_t and p_t relative to the responses in the RF model. In general equilibrium, the overall effect of the larger responses of w_t and p_t in the HF model is to dampen the decline in N_t that is directly due to the entry and exit margins (Figure 2.8, orange line). p_t is the real marginal cost faced by intermediate goods producers, so a larger decline in p_t will also be associated with a larger decline in inflation.

Figure 2.8: Difference in Response of Labour Demand Relative to RF model



Notes: Blue line is the response of labour demand to the path of prices obtained from the RF model (assuming perfect foresight about prices) minus the response of labour demand in the RF model; orange line is the difference between the responses of labour demand in the HF and RF models.

Why is there little additional endogenous amplification of the monetary policy shock due to the entry and exit margins in this model? This partly reflects the small responses of entry and exit rates to the shock. However, I show below that there is still little amplification in a version of the model that has been modified to target the responses of entry and exit rates estimated using the proxy SVAR in Section 2.2. The other reason for the small degree of amplification is that the firms that are induced to exit (or not enter) due to the shock tend to be lower-productivity firms, which are small. The entry and exit margins therefore contribute only a small additional reduction in labour demand above and beyond the reduction in labour demand due to firms operating at smaller scale. Additionally, as described above and illustrated in Figure 2.8, the role of entry and exit in amplifying the shock's effects on aggregate employment – which is quantitatively small to begin with – is dampened by the general-equilibrium responses of prices.

The entry and exit margins also serve to endogenously propagate the monetary policy shock. The state variable in the RF model is the monetary policy shock, so there is no endogenous state and no endogenous propagation of shocks (the pth-order autocorrelation of any variable in the RF model is simply ρ_m^p). In contrast, the measure of incumbent firms over idiosyncratic productivity is the endogenous state in the HF model, which is shaped by changes in entry and exit rates across the productivity distribution, so there is endogenous propagation of the monetary policy shock. However, given the similarity of responses in the HF and RF models, this endogenous propagation mechanism appears to be quantitatively unimportant. For example, the four-quarter autocorrelation of output is 0.064 in the HF model compared with 0.0625 in the RF model.

The endogenous response of aggregate TFP. Aggregating individual production firms' output yields an aggregate production function of the form $Y_t = A_t N_t^{\nu}$, where

$$A_t = \left[\Gamma_t \mathbb{E}_z\left(z^{\frac{1}{1-\nu}}\right)\right]^{1-\nu} \tag{2.17}$$

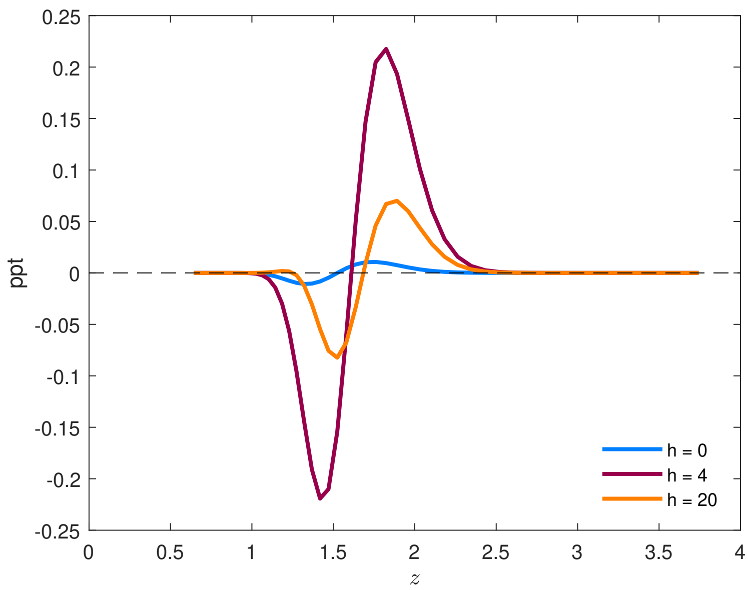
and $\mathbb{E}_z(.)$ is with respect to the distribution of firms, $\mu_t(z)/\Gamma_t$. The variable A_t is a natural measure of aggregate TFP. Changes in entry and exit rates due to a monetary policy shock induce changes in the measure of firms operating. If the responses of entry and exit rates differ across the distribution of firms, the distribution over productivity will also respond to the shock. Aggregate TFP in this model is therefore endogenous, which is a channel of monetary policy transmission that is absent in the textbook New Keynesian model.

Figure 2.6 shows that the shock causes a small, but highly persistent, decrease in aggregate TFP. Figure 2.9 plots the change in the distribution of firms at each level of productivity for different horizons after the realisation of the monetary policy shock. On impact, there is a decline in the measure of firms operating at all levels of productivity due to a decline in entry rates (operating firms who choose to exit following the shock only do so in the subsequent period). This decline tends to be larger for firms with lower productivity, so there is a small shift upwards in the productivity distribution, which increases one of the two components of aggregate TFP, $\mathbb{E}_z\left(z^{\frac{1}{1-v}}\right)$. However, the decline in Γ_t more than offsets this effect, so aggregate TFP falls slightly. Four quarters after the shock, the change in the distribution of firms over idiosyncratic productivity is more pronounced due to the increase in exit rates, which again tends to be larger for lower-productivity firms, but the decline in the measure of firms operating continues to outweigh this effect, so aggregate TFP is still lower. Subsequently, the distribution of firms only slowly reverts to its stationary distribution due to the persistent process for idiosyncratic productivity. The change in entry and exit rates has long-lasting effects on both the productivity distribution and the total measure of operating firms, so the change in aggregate TFP is extremely persistent (the four-quarter autocorrelation in A_t is around 0.95), albeit small in magnitude.[21]

What have we learned? Even though the monetary policy shock is endogenously amplified and propagated by changes in entry and exit rates and a resulting change in aggregate TFP, these channels of transmission appear to be quantitatively unim-

[21] Consistent with the model's prediction, Moran and Queralto (2018) estimate that aggregate TFP declines following a contractionary monetary policy shock using an SVAR in which the monetary policy shock is identified using a causal ordering. Adding measures of aggregate TFP from Fernald (2012) to the proxy SVAR from Section 2.2 also suggests that aggregate TFP may decline following a contractionary shock, although the 90 per cent confidence intervals for the response of utilisation-adjusted TFP include zero at all horizons.

Figure 2.9: Change in Distribution of Firms



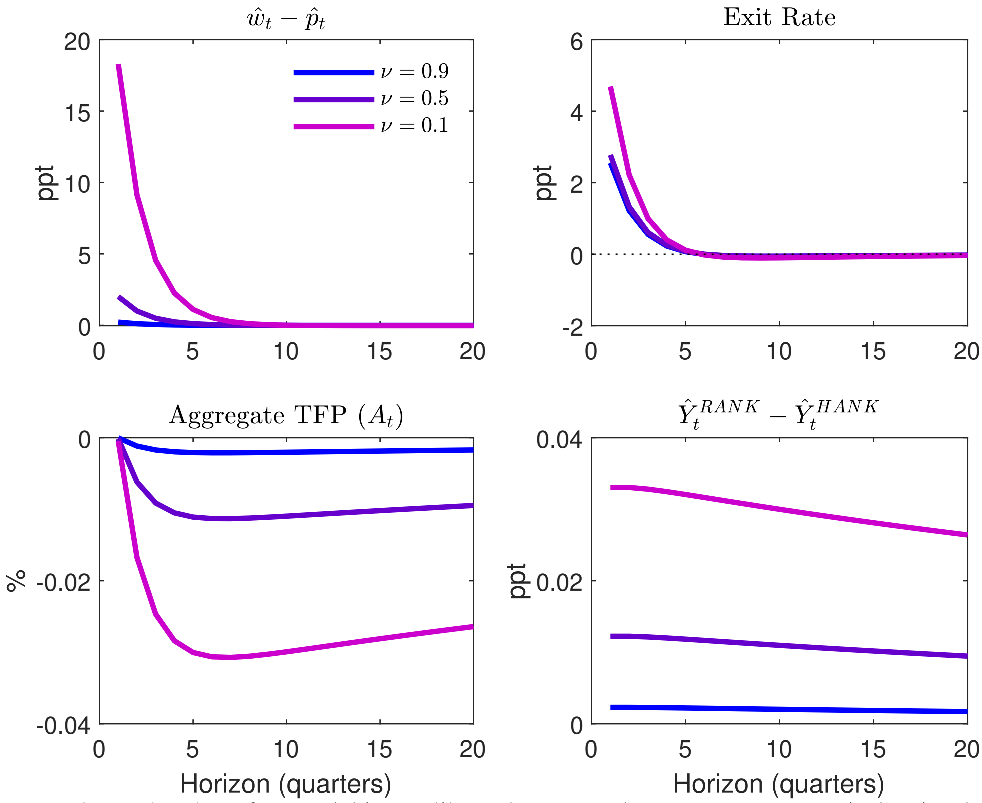
Notes: h is the number of quarters after the monetary policy shock; distribution of firms is $\mu_t(z)/\Gamma_t$.

portant in the sense that impulse responses from a comparable model without these channels are almost identical. A pertinent question is the extent to which this result depends on the calibration or on particular features of the model's structure. I explore this in the next section.

2.4.3 Robustness

Returns to scale and the co-movement of w_t and p_t. As discussed above, a key feature of the model's responses to the shock is the strong co-movement of w_t and p_t, which results in the responses of entry and exit rates being predominately driven by the direct effect of the change in the real interest rate. The strength of this co-movement will be affected by the degree of returns to scale, which is controlled by ν. With lower ν, a given decline in p_t will result in a smaller decline in labour demand than when ν is higher, because more-strongly decreasing returns to scale mean that a smaller reduction in employment is required to achieve the same reduction in marginal cost. In turn, the smaller decline in labour demand yields a smaller decline in w_t. A smaller decline in w_t for any decline in p_t means that entry and exit rates will respond more strongly to the shock. Smaller values of ν therefore generate larger changes in entry and exit rates, and larger differences between the responses of variables in the HF and RF models. However, even implausibly small values of ν do not generate noticeable differences between the responses of variables in the HF and RF models. For example, assuming $\nu = 0.1$ generates a large wedge between the responses of p_t and w_t, but the exit rate increases by only 4–5 basis points and the decline in output is only about 0.03 percentage points larger in the HF model than in the RF model (Figure 2.10).

Figure 2.10: Responses Under Alternative Values of ν



Notes: Under each value of ν, model is recalibrated to target the same moments as in Section 2.4.1; $\hat{x}_t$ is the log-deviation of the variable x_t from its value in the stationary equilibrium.

Sensitivity to other parameter values. Changing the parameters of the model one-at-a-time without recalibrating the model tends to have little impact on the qualitative results with the exception of the parameters characterising the distribution of fixed costs. This is because the responses of entry and exit rates to the shock depend partly on the shape of the distribution of fixed costs. For example, increasing μ_c or σ_c sufficiently can yield responses of entry and exit rates to the shock that are similar to those estimated in the data. This generates larger differences between the responses of variables in the HF and RF models, although these difference are still small (e.g., the difference in the responses of output is on the order of one-tenth of a percentage point) and come at the cost of an unrealistically high exit rate in the stationary equilibrium.

Matching the empirical responses of entry and exit rates. The impact responses of entry and exit rates in the model are smaller in magnitude than the peak responses estimated in the data. A natural question is whether a model with empirically realistic responses of entry and exit rates can generate meaningfully different impulse responses to the RF model. One way to increase the responsiveness of entry and exit rates to the shock without considerably altering the structure of the model or changing its stationary equilibrium is to allow the distribution of fixed costs to depend directly on the deviation of the real interest rate from its value in the stationary equilibrium. In particular, I suppose that $c \sim LN(\mu_c + \alpha_c(r_t - 1/\beta), \sigma_c)$ and $e \sim LN(\mu_c + \alpha_e(r_t - 1/\beta), \sigma_c)$, where α_c and α_e are calibrated such that the exit rate increases by 10 basis points and the entry rate decreases by 4.5 basis points in response to the shock. Even with these larger responses of entry and exit rates, the responses of macroeconomic variables are still very similar in the HF and RF models; for example, the difference in the impact response of output is only about one-tenth of a percentage point.

Alternative model structures. The similarity between the responses of variables in the HF and RF models also holds under several alternative assumptions about the structure of the model. These include the assumptions that: fixed costs are denominated in either units of the production good or labour (rather than the final good); potential entrants who pay the entry cost do not operate until the subsequent period; production firms are risk neutral; or there is free entry of firms instead of a fixed mass of potential entrants. It also holds in a version of the model that adds physical capital and other features of typical medium-scale New Keynesian models, including aggregate investment-adjustment costs, habits in consumption and sticky wages. These alterations of the model are described in more detail in Appendix A.3.

The reason why the responses of variables in the HF and RF models remain similar under these alternative specifications is that movements in p_t and the other prices that enter the production firms' decision problem continue to largely offset one another, so the responses of entry and exit rates are driven predominately by the direct effect of the real interest rate. Consequently, the responses remain quantitatively small. Moreover, the firms that are induced to exit (or not enter) still tend to be those with lower productivity (i.e., small firms), so responses along the extensive margin do not result in a large additional reduction in labour demand.

Firm-level frictions? As discussed above, part of the reason for the limited additional amplification of the monetary policy shock in the HF model relative to the RF model is that the firms that are induced to exit (or not enter) due to the shock tend to be low-productivity firms, which are small. Changes in entry and exit therefore result in little additional reduction in labour demand above and beyond the reduction in labour demand due to changes along the intensive margin, and there is little additional effect on equilibrium prices or aggregate quantities. Frictions that make larger firms' entry and exit decisions more sensitive to changes in prices should result in these margins playing a more important role in the transmission of monetary policy shocks, because any given change in entry or exit rates will result in a larger change in labour demand, and thus greater amplification of the shock. For example, the presence of labour-adjustment costs (as in Hopenhayn and Rogerson (1993)) would reduce the ability of operating firms to adjust employment along the intensive margin in response to changes in prices and could result in more exit among larger firms. Models with investment-adjustment costs (as in Clementi and Palazzo (2016)) or financial frictions (as in Ottonello and Winberry (2020)) may also be able to generate more-prominent roles for entry and exit in amplifying and propagating monetary policy shocks if these frictions make the entry and exit decisions of larger firms (in terms of employment or capital) more sensitive to changes in prices. I leave exploration of these avenues for future work.[22]

[22] In assessing the role of entry and exit in amplifying the shock in a model with firm-level frictions, one would want to compare responses against comparable models. The relevant model may be a representative-firm model with the given friction, or a heterogeneous-firm model with the given friction but without endogenous entry and exit. A version of my model without endogenous entry and exit would be equivalent to a representative-firm model, since $\mu_t(z)$ would be constant.

2.5 Conclusion

A New Keynesian extension of the standard firm-dynamics model is able to generate qualitatively similar responses of firm entry and exit rates to a monetary policy shock as those estimated using a proxy SVAR, in the sense that firms are more likely to exit and less likely to enter following a contractionary shock. In the model, the decrease in entry and the increase in exit tend to be larger for lower-productivity firms, which shifts the distribution of idiosyncratic TFP upwards. A decline in the measure of operating firms due to the decrease in entry and the increase in exit more than offsets the upward shift in the productivity distribution and results in aggregate TFP falling. However, the decline in aggregate TFP is quantitatively small and the responses of macroeconomic variables in the model are essentially indistinguishable from those in a comparable representative-firm model. Accounting for firm dynamics therefore appears to have little effect on predictions about the effects of monetary policy relative to those from a representative-firm model. This also holds under different parameter values and under alternative assumptions about the structure of the model.

I have deliberately abstracted from firm-level frictions to highlight the role (or lack thereof) of firm dynamics in the transmission of monetary policy in a simple New Keynesian firm-dynamics model. Future work could explore the role of entry and exit in monetary policy transmission using models with firm-level frictions. Frictions that result in larger firms' entry and exit behaviour being more sensitive to changes in prices are likely to result in these margins playing a more important role in the transmission of monetary policy shocks. Firm-level evidence on the entry and exit behaviour of firms following a monetary policy shock may be useful in disciplining the responses of entry and exit across the distribution of firms and in assessing the relevance of alternative frictions.

Chapter 3

Algorithms for Inference in SVARs Identified with Sign and Zero Restrictions

3.1 Introduction

Structural vector autoregressions (SVARs) are used in macroeconomics to estimate the dynamic causal effects of structural shocks. Parameters in these models have traditionally been point-identified using zero restrictions on the SVAR's structural parameters. However, it has become increasingly common to set-identify parameters using sign restrictions and/or a set of zero restrictions that are insufficient to achieve point-identification.[1] Inference in set-identified SVARs has typically been carried out via Bayesian methods that rely on rejection sampling.[2] For example, when there are both sign and zero restrictions, the algorithms in Arias et al. (2018) involve drawing parameter values satisfying the zero restrictions and discarding them if they do not satisfy the sign restrictions. A drawback of this approach is that it may be computationally demanding when the sign restrictions considerably truncate the identified set given the zero restrictions, because many draws of the parameters satisfying the zero restrictions may be required to obtain a sufficient

[1]For an overview of identification in SVARs, see Kilian and Lütkepohl (2017) or Stock and Watson (2016).

[2]For examples of frequentist approaches to inference in set-identified SVARs, see Gafarov et al. (2018) and Granziera et al. (2018).

number of draws that additionally satisfy the sign restrictions. To address this problem, this paper develops algorithms to facilitate Bayesian inference in SVARs that are set-identified using a combination of sign and zero restrictions.

It is convenient to parameterize set-identified SVARs in terms of the VAR's reduced-form parameters and an orthonormal matrix so that sign and zero restrictions can be expressed as restrictions on this matrix. I focus on the case where sign and zero restrictions linearly constrain a single column of the orthonormal matrix, which I denote by $\mathbf{q} \in \mathbb{R}^n$, where n is the dimension of the VAR. I also discuss extensions to the case where there are restrictions on multiple columns of the orthonormal matrix. The algorithms I develop are compatible with a wide range of sign and zero restrictions. These restrictions include sign restrictions on impulse responses (e.g. Uhlig, 2005), bounds on elasticities (e.g. Kilian and Murphy, 2012) and shape restrictions (e.g. Amir-Ahmadi and Drautzburg, 2021), as well as sign and zero restrictions on the structural parameters themselves. Other compatible zero restrictions include 'short-run' restrictions on impact impulse responses, as in Christiano et al. (1999) or Sims (1980), 'long-run' restrictions, as in Blanchard and Quah (1989), and restrictions arising from external instruments or 'proxies', as in Mertens and Ravn (2013) and Stock and Watson (2018). The algorithms can also accommodate certain types of 'narrative restrictions', including restrictions on the sign of a structural shock in particular periods (e.g. Antolín-Díaz and Rubio-Ramírez, 2018) or the timing of its maximum realization (e.g. Giacomini et al., 2021).

The algorithms developed in this paper build on those proposed in Amir-Ahmadi and Drautzburg (2021) (AD21), which are applicable when there are sign restrictions only. AD21 show that the problem of determining whether the identified set is nonempty can be cast as a linear program, which can be solved rapidly using standard software. They also propose a Gibbs sampler that draws $\mathbf{q}$ from a uniform distribution conditional on sign restrictions. Importantly, both algorithms avoid rejection sampling and so may be more computationally efficient than existing algorithms when the sign restrictions substantially truncate the identified set given the zero restrictions.[3]

To extend the algorithms in AD21 to allow for zero restrictions, I show how a system of sign and zero restrictions in $\mathbb{R}^n$ can be expressed as an equivalent system of sign restrictions in a lower-dimensional space. The algorithms in AD21 are applicable to the transformed system of sign restrictions and, in conjunction with a simple transformation, can be used to obtain values of the parameters satisfying the original identifying restrictions. Specifically, an algorithm determines whether the identified set is nonempty by solving a linear program and, if so, generates a value of $\mathbf{q}$ satisfying the identifying restrictions. This value of $\mathbf{q}$ can be used to initialize a Gibbs sampler that draws from a uniform distribution over the identified set for $\mathbf{q}$. Additionally, it can be used to initialize a gradient-based numerical optimization routine whose aim is to compute the bounds of the identified set for a scalar parameter of interest, which is useful in the context of prior-robust Bayesian inference (e.g. Giacomini and Kitagawa, 2021).

I illustrate the algorithms using the empirical application in Arias et al. (2019) (ACR19). They estimate the effects of monetary policy shocks in the United States by imposing sign and zero restrictions on the monetary policy reaction function. I augment these restrictions with the sign restrictions on impulse responses considered in Uhlig (2005), and explore the accuracy and computational efficiency of my algorithms relative to alternatives. My algorithms are particularly useful when a large number of sign restrictions appreciably truncate the identified set given the zero restrictions. These algorithms should therefore facilitate the use of rich sets of sign restrictions alongside zero restrictions.

As an additional illustration of the utility of the algorithms, I impose a restriction on the timing of the maximum realization of the monetary policy shock considered in Giacomini et al. (2021); in addition to the restrictions from ACR19 and Uhlig (2005), I impose that the monetary policy shock in October 1979 – the month in which Paul Volcker dramatically and unexpectedly raised the federal funds rate

[3] Rubio-Ramírez et al. (2010) describe an algorithm for drawing from a uniform distribution over the space of orthonormal matrices conditional on sign restrictions. Arias et al. (2018) extend this algorithm to allow for zero restrictions. Both algorithms use rejection sampling to impose that the draws satisfy the sign restrictions.

– was the largest positive realization of the shock in the sample. This restriction generates as many sign restrictions as observations in the sample (around 500 in this application) and tends to truncate the identified set considerably, which makes existing algorithms extremely computationally burdensome. Under this restriction, output falls with high posterior probability following a positive monetary policy shock. However, the identified set is empty in around 95 per cent of draws from the posterior of the reduced-form parameters, which suggests the restriction is inconsistent with the data.

Outline. The remainder of the paper is structured as follows. Section 3.2 outlines the SVAR framework and describes the identifying restrictions considered. Section 3.3 shows how a system of sign and zero restrictions can be expressed as an equivalent system of sign restrictions in a lower-dimensional space and explains how algorithms used to conduct inference in sign-restricted SVARs can consequently be extended to the case of zero restrictions. Section 3.4 describes how to numerically implement the algorithms. Section 3.5 explores the accuracy and efficiency of the algorithms relative to existing alternatives using the model in ACR19 augmented with additional identifying restrictions. Section 3.6 discusses extending the algorithms to allow for restrictions on additional columns of the orthonormal matrix. Section 3.7 concludes. Proofs are relegated to the Appendix.

Generic notation. For a matrix $\mathbf{X}$, $\text{vec}(\mathbf{X})$ is the vectorization of $\mathbf{X}$. When $\mathbf{X}$ is symmetric, $\text{vech}(\mathbf{X})$ is the half-vectorization of $\mathbf{X}$, which stacks the elements of $\mathbf{X}$ lying on or below the diagonal into a vector. $\mathbf{e}_{i,n}$ is the ith column of the $n \times n$ identity matrix, $\mathbf{I}_n$. $\mathbf{0}_{m\times n}$ is an $m \times n$ matrix of zeros. $\|.\|$ is the Euclidean norm. $\mathbb{S}^{n-1}$ is the unit sphere in $\mathbb{R}^n$ (i.e. the set $\{\mathbf{q} \in \mathbb{R}^n : \mathbf{q}'\mathbf{q} = 1\}$).

3.2 Framework

3.2.1 SVAR

Let $\mathbf{y}_t$ be an $n \times 1$ vector of endogenous variables following the SVAR(p) process:

$$\mathbf{A}_0\mathbf{y}_t = \sum_{l=1}^{p} \mathbf{A}_l\mathbf{y}_{t-l} + \boldsymbol{\varepsilon}_t, \quad t = 1,...,T, \tag{3.1}$$

where $\mathbf{A}_0$ is invertible and $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$ are structural shocks. The diagonal elements of $\mathbf{A}_0$ are normalized to be positive, which is a normalization on the signs of the structural shocks. Exogenous regressors (such as a constant) are omitted for simplicity of exposition, but these are straightforward to include. Letting $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$ and $\mathbf{A}_+ = (\mathbf{A}_1, \ldots, \mathbf{A}_p)$, rewrite the SVAR($p$) as

$$\mathbf{A}_0\mathbf{y}_t = \mathbf{A}_+\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T. \tag{3.2}$$

$(\mathbf{A}_0, \mathbf{A}_+)$ are the structural parameters. The reduced-form VAR(p) representation is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u}_t, \quad t = 1, ..., T, \tag{3.3}$$

where $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_p)$, $\mathbf{B}_l = \mathbf{A}_0^{-1}\mathbf{A}_l$ for $l = 1, \ldots, p$, and $\mathbf{u}_t = \mathbf{A}_0^{-1}\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{n\times 1}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$. $\boldsymbol{\phi} = (\text{vec}(\mathbf{B})', \text{vech}(\boldsymbol{\Sigma})')' \in \boldsymbol{\Phi}$ are the reduced-form parameters. The SVAR's orthogonal reduced form is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \boldsymbol{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{3.4}$$

where $\boldsymbol{\Sigma}_{tr}$ is the lower-triangular Cholesky factor of $\boldsymbol{\Sigma}$ (i.e. $\boldsymbol{\Sigma}_{tr}\boldsymbol{\Sigma}'_{tr} = \boldsymbol{\Sigma}$) with non-negative diagonal elements and $\mathbf{Q}$ is an $n \times n$ orthonormal matrix with jth column $\mathbf{q}_j$. Let $\mathscr{O}(n)$ denote the space of all $n \times n$ orthonormal matrices.

Impulse responses are typically the parameters of interest in analyses using SVARs. The horizon-h impulse response of the ith variable to the jth shock is

$$\eta_{i,j,h} \equiv \eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{q}_j) = \mathbf{e}'_{i,n}\mathbf{C}_h\boldsymbol{\Sigma}_{tr}\mathbf{q}_j, \tag{3.5}$$

where $\mathbf{C}_h$ is defined recursively by $\mathbf{C}_h = \sum_{l=1}^{\min\{h,p\}} \mathbf{B}_l\mathbf{C}_{h-l}$ for $h \geq 1$ with $\mathbf{C}_0 = \mathbf{I}_n$.

3.2.2 Identifying restrictions

Consider the case where there are linear sign and zero restrictions constraining $\mathbf{q}_1$ only (extensions to this case are discussed in Section 3.6). Let $\mathbf{F}(\boldsymbol{\phi})$ be the $r \times n$ matrix whose rows represent the coefficients of r zero restrictions, so $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 =$

$\mathbf{0}_{r\times 1}$. For example, zero restrictions on the first row of $\mathbf{A}_0$ take the form $\mathbf{e}'_{1,n}\mathbf{A}_0\mathbf{e}_{i,n} = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{i,n})'\mathbf{q}_1 = 0$, zero restrictions on impact responses to the first shock take the form $\mathbf{e}'_{i,n}\mathbf{A}_0^{-1}\mathbf{e}_{1,n} = \mathbf{e}'_{i,n}\mathbf{\Sigma}_{tr}\mathbf{q}_1 = 0$ and long-run restrictions on cumulative impulse responses to the first shock take the form $\mathbf{e}'_{i,n}(\mathbf{I}_n - \sum_{l=1}^{p}\mathbf{B}_l)^{-1}\mathbf{\Sigma}_{tr}\mathbf{q}_1 = 0$. Assume $0 < r < n-1$, which implies $\mathbf{q}_1$ is set-identified (Rubio-Ramírez et al., 2010), and assume $\text{rank}(\mathbf{F}(\boldsymbol{\phi})) = r$.

Similarly, let $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$ represent a set of s sign restrictions, which includes the sign normalization $\mathbf{e}'_{1,n}\mathbf{A}_0\mathbf{e}_{1,n} = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\mathbf{q}_1 \geq 0$. $\mathbf{S}(\boldsymbol{\phi})$ may include restrictions on impulse responses to a standard-deviation shock, ratios of these impulse responses (e.g. elasticity and shape restrictions) and/or elements of the first row of $\mathbf{A}_0$. For example, a bound on the impact impulse response of the ith variable to a shock in the first variable that raises the first variable by one unit is $(\mathbf{e}'_{i,n}\mathbf{\Sigma}_{tr}\mathbf{q}_1)/(\mathbf{e}'_{1,n}\mathbf{\Sigma}_{tr}\mathbf{q}_1) \geq \lambda$, where λ is a known scalar. This restriction can be expressed as $(\mathbf{e}'_{i,n} - \lambda\mathbf{e}'_{1,n})\mathbf{\Sigma}_{tr}\mathbf{q}_1 \geq 0$. An example of a shape restriction is that the horizon-h impulse response of the first variable to the first shock is weakly greater than the horizon-l response, which requires that $\mathbf{e}'_{i,n}\mathbf{C}_h\mathbf{\Sigma}_{tr}\mathbf{q}_1 \geq \mathbf{e}'_{i,n}\mathbf{C}_l\mathbf{\Sigma}_{tr}\mathbf{q}_1$ or $\mathbf{e}'_{i,n}(\mathbf{C}_h - \mathbf{C}_l)\mathbf{\Sigma}_{tr}\mathbf{q}_1 \geq 0$. $\mathbf{S}(\boldsymbol{\phi})$ may also include particular types of narrative restrictions, including restrictions on the sign or relative magnitude of the first shock in particular periods. For example, the restriction that the first shock is nonnegative in period k is $\mathbf{e}'_{1,n}\mathbf{A}_0\mathbf{u}_k = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{u}_k)'\mathbf{q}_1 \geq 0$. The restriction that the first shock in period k is larger than the first shock in period m is $\mathbf{e}'_{1,n}\mathbf{A}_0\mathbf{u}_k \geq \mathbf{e}'_{1,n}\mathbf{A}_0\mathbf{u}_m$, which is equivalent to $(\mathbf{\Sigma}_{tr}^{-1}(\mathbf{u}_k - \mathbf{u}_m))'\mathbf{q}_1 \geq 0$.[4]

Given a set of identifying restrictions, the identified set for $\mathbf{q}_1$ collects observationally equivalent parameter values and is defined as

$$\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S}) = \left\{\mathbf{q}_1 \in \mathbb{S}^{n-1} : \mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}, \mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}\right\}. \tag{3.6}$$

This set has a geometric interpretation. The zero restrictions $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ restrict $\mathbf{q}_1$ to lie in an $(n-r)$-dimensional hyperplane, while the sign restrictions $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq$

[4] Under narrative restrictions, $\mathbf{S}(\boldsymbol{\phi})$ is also a function of the data in particular periods through the reduced-form innovations, but I leave this potential dependence implicit. See Antolín-Díaz and Rubio-Ramírez (2018) or Giacomini et al. (2021) for further details on narrative restrictions.

$\mathbf{0}_{s\times 1}$ restrict $\mathbf{q}_1$ to lie within the intersection of s half-spaces. Since $\mathbf{q}_1$ is a column of an orthonormal matrix, it must have unit length, so it lies on the unit sphere in $\mathbb{R}^n$. The identified set for $\mathbf{q}_1$ is the intersection of these spaces, which may be empty at particular values of $\boldsymbol{\phi}$. The identified set for an impulse response $\eta_{i,j,h}$ is

$$IS_\eta(\boldsymbol{\phi}|\mathbf{F},\mathbf{S}) = \left\{\eta_{i,j,h}(\boldsymbol{\phi},\mathbf{q}_1) : \mathbf{q}_1 \in \mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})\right\}. \tag{3.7}$$

3.3 Transforming the System of Identifying Restrictions

This section shows that the system of equality and inequality restrictions in $\mathbb{R}^n$ can be expressed as an equivalent system of inequality restrictions in $\mathbb{R}^{n-r}$. Subsequently, I explain how the algorithms proposed in AD21 for the case of sign restrictions can be used to check whether $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ is nonempty and, if so, to generate a value of $\mathbf{q}_1$ satisfying the identifying restrictions. Additionally, the Gibbs sampler developed in AD21 can be extended to randomly sample $\mathbf{q}_1$ from a uniform distribution over $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$.

Let $N(\mathbf{F}(\boldsymbol{\phi}))$ denote an orthonormal basis for the null space of $\mathbf{F}(\boldsymbol{\phi})$, which spans the hyperplane $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$. Under the assumption $\text{rank}(\mathbf{F}(\boldsymbol{\phi})) = r$, the rank-nullity theorem implies $N(\mathbf{F}(\boldsymbol{\phi}))$ is of dimension $n-r$. The null space of $N(\mathbf{F}(\boldsymbol{\phi}))'$ is then of dimension r and the columns of the matrix $\mathbf{K} = (N(\mathbf{F}(\boldsymbol{\phi})), N(N(\mathbf{F}(\boldsymbol{\phi}))'))$ form an orthonormal basis for $\mathbb{R}^n$.[5] The matrix that transforms from this basis into the standard basis is $\mathbf{K}^{-1}$. In the new basis, the coefficients in the zero and sign restrictions are, respectively, $\tilde{\mathbf{F}}(\boldsymbol{\phi}) = (\mathbf{K}^{-1}\mathbf{F}(\boldsymbol{\phi})')'$ and $\tilde{\mathbf{S}}(\boldsymbol{\phi}) = (\mathbf{K}^{-1}\mathbf{S}(\boldsymbol{\phi})')'$. After applying this change of basis, the hyperplane generated by the zero restrictions coincides with the hyperplane spanned by the first $n-r$ basis vectors (i.e. the first $n-r$ column vectors of $\mathbf{I}_n$). Any vector lying in this hyperplane will therefore have its last r elements equal to zero.

After the change of basis, the projection of the ith row of $\tilde{\mathbf{S}}(\boldsymbol{\phi})$, $\tilde{\mathbf{S}}_i(\boldsymbol{\phi})$, onto the

[5] I leave the dependence of $\mathbf{K}$ on $\mathbf{F}(\boldsymbol{\phi})$ implicit.

hyperplane generated by the zero restrictions is[6]

$$\bar{\mathbf{S}}_i(\boldsymbol{\phi})' = (\mathbf{I}_n - \tilde{\mathbf{F}}(\boldsymbol{\phi})'(\tilde{\mathbf{F}}(\boldsymbol{\phi})\tilde{\mathbf{F}}(\boldsymbol{\phi})')^{-1}\tilde{\mathbf{F}}(\boldsymbol{\phi}))\tilde{\mathbf{S}}_i(\boldsymbol{\phi})'. \tag{3.8}$$

Let $\mathbf{M} = (\mathbf{I}_{n-r}, \mathbf{0}_{(n-r)\times r})$ be the $(n-r)\times n$ matrix such that $\mathbf{Mx}$ drops the last r elements of the $n \times 1$ vector $\mathbf{x}$ and let $\bar{\mathbf{S}}(\boldsymbol{\phi}) = (\mathbf{M}\bar{\mathbf{S}}_i(\boldsymbol{\phi})', \ldots, \mathbf{M}\bar{\mathbf{S}}_s(\boldsymbol{\phi})')'$. The end result of these transformations is that the sign and zero restrictions in $\mathbb{R}^n$ have been replaced with an equivalent system of sign restrictions $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$ in $\mathbb{R}^{n-r}$. This claim is formalized in the following proposition.

Proposition 3.3.1. *Let* $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ *be a system of* r *zero restrictions with* $\text{rank}(\mathbf{F}(\boldsymbol{\phi})) = r$ *and let* $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$ *be a system of* s *sign restrictions.*
(a) If $\mathbf{q}_1 \in \mathbb{R}^n$ *satisfies* $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ *and* $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$*, then* $\bar{\mathbf{q}}_1 = \mathbf{M}\mathbf{K}^{-1}\mathbf{q}_1 \in \mathbb{R}^{n-r}$ *satisfies* $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$.
(b) If $\bar{\mathbf{q}}_1 \in \mathbb{R}^{n-r}$ *satisfies* $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$*, then* $\mathbf{q}_1 = \mathbf{K}\mathbf{M}'\bar{\mathbf{q}}_1 \in \mathbb{R}^n$ *satisfies* $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ *and* $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$.

This proposition implies the following corollary relating (non)emptiness of the set $\bar{\mathscr{Q}}_1(\boldsymbol{\phi}|\bar{\mathbf{S}}) = \left\{\bar{\mathbf{q}}_1 \in \mathbb{S}^{n-r-1} : \bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}\right\}$ to (non)emptiness of the set $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$.

Corollary 3.3.1. $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ *is nonempty if and only if* $\bar{\mathscr{Q}}_1(\boldsymbol{\phi}|\bar{\mathbf{S}})$ *is nonempty.*

Based on the results in AD21, $\bar{\mathscr{Q}}_1(\boldsymbol{\phi}|\bar{\mathbf{S}})$ is nonempty if the largest ball that can be inscribed within the intersection of the s half-spaces generated by the inequality restrictions $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$ and the unit $(n-r)$-cube has positive radius. The problem of finding the radius and 'Chebyshev' centre of this ball can be formulated as a linear program, which can be solved efficiently (e.g. Boyd and Vandenberghe, 2004). If the ball has positive radius with centre $\mathbf{c} \in \mathbb{R}^{n-r}$, then $\bar{\mathbf{q}}_1^{(0)} = \mathbf{c}/\|\mathbf{c}\|$ satisfies $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1^{(0)} \geq \mathbf{0}_{s\times 1}$ and lies in $\mathbb{S}^{n-r-1}$. By Proposition 3.3.1(ii), $\mathbf{q}_1^{(0)} = \mathbf{K}\mathbf{M}'\bar{\mathbf{q}}_1^{(0)}$ satisfies the original set of identifying restrictions and lies in $\mathbb{S}^{n-1}$.

[6]Since the hyperplane generated by the zero restrictions coincides with the hyperplane spanned by the first $n-r$ basis vectors, this is equivalent to projecting onto the linear subspace spanned by the first $n-r$ basis vectors via $\bar{\mathbf{S}}_i(\boldsymbol{\phi})' = (\mathbf{I}_n - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}')\tilde{\mathbf{S}}_i(\boldsymbol{\phi})'$, where $\mathbf{B} = (\mathbf{0}_{r\times(n-r)}, \mathbf{I}_r)'$ contains the last r basis vectors.

The Gibbs sampler in AD21 can be used to obtain a sequence of draws of $\bar{\mathbf{q}}_1$ from a uniform distribution over $\bar{\mathscr{Q}}_1(\boldsymbol{\phi}|\bar{\mathbf{S}})$ using $\bar{\mathbf{q}}_1^{(0)}$ to initialize the sampler.[7] Let $\bar{\mathbf{q}}_1^{(k)}$ represent the kth draw. If $\bar{\mathbf{q}}_1^{(k)}$ is uniformly distributed over $\bar{\mathscr{Q}}_1(\boldsymbol{\phi}|\bar{\mathbf{S}})$, then $\mathbf{M}'\bar{\mathbf{q}}_1^{(k)}$ is uniformly distributed over

$$\left\{\mathbf{M}'\bar{\mathbf{q}}_1 \in \mathbb{S}^{n-1} : \tilde{\mathbf{F}}(\boldsymbol{\phi})\mathbf{M}'\bar{\mathbf{q}}_1 = \mathbf{0}_{r\times 1}, (\mathbf{M}'\bar{\mathbf{S}}(\boldsymbol{\phi})')'\mathbf{M}'\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}\right\}. \tag{3.9}$$

Since $\mathbf{K}$ is an orthonormal matrix, $\mathbf{q}_1^{(k)} = \mathbf{K}\mathbf{M}'\bar{\mathbf{q}}_1^{(k)}$ is also uniformly distributed. Applying this transformation to each draw $\bar{\mathbf{q}}_1^{(k)}$ therefore yields draws $\mathbf{q}_1^{(k)}$ that are uniformly distributed over $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$. These transformed draws can be used when conducting Bayesian inference under a conditionally uniform prior for $\mathbf{q}_1$ given $\boldsymbol{\phi}$.[8]

To provide some geometric intuition, it is useful to consider the case where $n = 3$ and there is one zero restriction. The set $\bar{\mathscr{Q}}_1(\boldsymbol{\phi}|\bar{\mathbf{S}})$ (when it is nonempty) is an arc of the unit circle in $\mathbb{R}^2$. The Gibbs sampler from AD21 generates draws from a uniform distribution over this arc. Applying the transformation $\mathbf{M}'\bar{\mathbf{q}}_1^{(k)}$ to the draws embeds the draws on this arc as draws on an arc of the unit sphere in $\mathbb{R}^3$, where the arc lies within the plane perpendicular to the z axis. Since $\mathbf{K}$ is orthonormal, left-multiplication by $\mathbf{K}$ rotates the draws about a particular axis of rotation. The rotation preserves the distribution of the draws on the arc, which now lies within the plane perpendicular to the vector $\mathbf{F}(\boldsymbol{\phi})'$.

3.4 Numerical Implementation

This section describes numerical algorithms to facilitate inference in SVARs identified using sign and zero restrictions. Algorithm 3.4.1 determines whether the identified set is nonempty and, if so, generates a value of $\mathbf{q}_1$ satisfying the identifying restrictions. Algorithm 3.4.2 generates draws of $\mathbf{q}_1$ that are uniformly distributed

[7] The Gibbs sampler in AD21 builds on a Gibbs sampler developed by Li and Ghosh (2015) for sampling from a multivariate normal distribution truncated by linear inequality restrictions.

[8] The accept-reject sampler proposed in Arias et al. (2018) involves rejecting joint draws of $(\boldsymbol{\phi},\mathbf{Q})$ that violate the sign restrictions. In contrast, imposing a conditionally uniform prior requires obtaining a single draw of $\mathbf{Q}$ (or $\mathbf{q}_1$) that satisfies the identifying restrictions at each draw of the reduced-form parameters such that the identified set is nonempty. See Uhlig (2017) for a discussion of this point.

over $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ via Gibbs sampling. The algorithms operate given a value of $\boldsymbol{\phi}$ and can be embedded within a posterior sampler for these parameters, in which case the assumption $\text{rank}(\mathbf{F}(\boldsymbol{\phi})) = r$ needs to hold $\boldsymbol{\phi}$-almost surely. For convenience, I suppress dependence on $\boldsymbol{\phi}$ in the descriptions of the algorithms below.

Algorithm 3.4.1. Determining whether $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ is empty. Let $\mathbf{F}\mathbf{q}_1 \geq \mathbf{0}_{r\times 1}$ be the set of zero restrictions and let $\mathbf{S}\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$ be the set of sign restrictions (including the sign normalization) given $\boldsymbol{\phi}$.

- ***Step 1.*** Compute the change-of-basis matrix $\mathbf{K} = (N(\mathbf{F}), N(N(\mathbf{F})'))$ and transform the coefficient vectors of the sign and zero restrictions into the new basis via $\tilde{\mathbf{S}} = (\mathbf{K}^{-1}\mathbf{S}')'$ and $\tilde{\mathbf{F}} = (\mathbf{K}^{-1}\mathbf{F}')'$.[9]

- ***Step 2.*** Project the coefficient vectors of the sign restrictions in the new basis onto the linear subspace spanned by the rows of $\tilde{\mathbf{F}}$ and drop the last r elements of the resulting vectors. The transformed matrix of coefficients is $\bar{\mathbf{S}} = (\mathbf{M}(\mathbf{I}_n - \tilde{\mathbf{F}}'(\tilde{\mathbf{F}}\tilde{\mathbf{F}}')^{-1}\tilde{\mathbf{F}})\tilde{\mathbf{S}}')'$, where $\mathbf{M} = (\mathbf{I}_{n-r}, \mathbf{0}_{(n-r)\times r})$.

- ***Step 3.*** Solve for the Chebyshev centre $\mathbf{c} = (c_1, \ldots, c_{n-r})'$ and radius R of the set $\{\bar{\mathbf{q}}_1 \in \mathbb{R}^{n-r} : \bar{\mathbf{S}}\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}, |\bar{q}_{1,i}| \leq 1, i = 1, \ldots, n-r\}$, where $\bar{q}_{1,i}$ is the ith element of $\bar{\mathbf{q}}_1$, by solving the linear program:

$$\max_{\{R\geq 0,\mathbf{c}\}} R$$

  subject to

$$\begin{aligned} \mathbf{e}_{k,s}'\bar{\mathbf{S}}\mathbf{c} + R\|\mathbf{e}_{k,s}'\bar{\mathbf{S}}\| \geq 0, &\quad k = 1, \ldots, s, \\ c_i + R \leq 1, &\quad i = 1, \ldots, n-r, \\ c_i - R \geq -1, &\quad i = 1, \ldots, n-r. \end{aligned}$$

- ***Step 4.*** If $R > 0$, conclude $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ is nonempty and compute $\bar{\mathbf{q}}_1^{(0)} = \mathbf{c}/\|\mathbf{c}\|$.

[9] In the MATLAB code accompanying the paper, I implement this step using MATLAB's 'null' function, which uses the singular value decomposition to compute an orthonormal basis for the null space.

Otherwise, conclude $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ is empty.

If interest is in computing the bounds of the identified set for a scalar function of $\boldsymbol{\phi}$ and $\mathbf{q}_1$, such as $\eta_{i,j,h}$, $\mathbf{q}_1^{(0)} = \mathbf{KM}'\bar{\mathbf{q}}_1^{(0)}$ is a feasible value of $\mathbf{q}_1$ satisfying the identifying restrictions and can be used to initialize a gradient-based optimization algorithm. This is relevant in the context of conducting prior-robust Bayesian inference, as in Giacomini and Kitagawa (2021).

If interest is in obtaining uniformly distributed draws over $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$, $\bar{\mathbf{q}}_1^{(0)}$ can be used to initialize the following Gibbs sampler.

Algorithm 3.4.2. Gibbs sampler for $\mathbf{q}_1$. Assume the output of Algorithm 3.4.1 is available and $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ is nonempty. Initialize the algorithm at $\mathbf{z}^{(0)} = \bar{\mathbf{q}}_1^{(0)}$ and let L be the desired number of draws of $\mathbf{q}_1$. For $k = 1,\ldots,L$, iterate on the following steps:

- ***Step 1.*** Let $\bar{\mathbf{S}}_{j,v:w}$ be elements $v, v+1,\ldots,w-1,w$ of the jth row of $\bar{\mathbf{S}}$ and let $\mathbf{z}_{v:w}^{(k)}$ be elements $v, v+1,\ldots,w-1,w$ of $\mathbf{z}^{(k)}$. For $i = 1,\ldots,n-r$, draw $z_i^{(k)}$ from the truncated standard normal distribution with lower bound $l_i^{(k)}$ and upper bound $u_i^{(k)}$, where

$$
\begin{aligned}
l_i^{(k)} &= \max_{j=1,\ldots,n-r:\bar{\mathbf{S}}_{j,i}>0} \left\{ -\frac{\bar{\mathbf{S}}_{j,1:(i-1)}\mathbf{z}_{1:(i-1)}^{(k)} + \bar{\mathbf{S}}_{j,(i+1)}\mathbf{z}_{(i+1):(n-r)}^{(k-1)}}{\bar{\mathbf{S}}_{j,i}} \right\}, \\
u_i^{(k)} &= \min_{j=1,\ldots,n-r:\bar{\mathbf{S}}_{j,i}<0} \left\{ -\frac{\bar{\mathbf{S}}_{j,1:(i-1)}\mathbf{z}_{1:(i-1)}^{(k)} + \bar{\mathbf{S}}_{j,(i+1):(n-r)}\mathbf{z}_{(i+1):(n-r)}^{(k-1)}}{\bar{\mathbf{S}}_{j,i}} \right\}
\end{aligned}
$$

  with $l_i^{(k)} = -\infty$ ($u_i^{(k)} = \infty$) if $\bar{\mathbf{S}}_{j,i} > 0$ ($\bar{\mathbf{S}}_{j,i} < 0$) does not hold for any j.

- ***Step 2.*** Compute $\bar{\mathbf{q}}_1^{(k)} = \mathbf{z}^{(k)}/\|\mathbf{z}^{(k)}\|$ and $\mathbf{q}_1^{(k)} = \mathbf{KM}'\bar{\mathbf{q}}_1^{(k)}$.

After discarding an appropriate number of initial draws, $\mathbf{q}_1^{(k)}$ can be considered as dependent draws from the uniform distribution over $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$. To obtain (approximately) independent draws, keep only every fth draw, where f is chosen such that so the retained draws are serially uncorrelated. To implement Step 1 in practice, I follow AD21 by drawing from the truncated standard normal distribution using the inverse cumulative distribution function (CDF) method. Letting $u \sim U(0,1)$, $\Phi^{-1}(u(\Phi(b)-\Phi(a))+\Phi(a))$ is a truncated standard normal random variable with lower truncation point a and upper truncation point b, where $\Phi(.)$ is the CDF of a standard normal random variable and $\Phi^{-1}(.)$ is the inverse CDF.

3.5 Empirical Illustration

This section applies the new algorithms in an empirical setting and compares their performance against existing alternatives.[10] The empirical application is from ACR19, who estimate the effects of monetary policy shocks in the United States.

Reduced-form VAR. The model's endogenous variables are real GDP (GDP_t), the GDP deflator ($GDPDEF_t$), a commodity price index (COM_t), total reserves (TR_t), nonborrowed reserves (NBR_t) (all in natural logarithms) and the federal funds rate (FFR_t). The data are monthly and run from January 1965 to June 2007. The VAR includes 12 lags.

I follow ACR19 by assuming a diffuse normal-inverse-Wishart prior over the reduced-form parameters. The posterior for the reduced-form parameters is then also a normal-inverse-Wishart distribution, from which it is straightforward to obtain independent draws (e.g. Del Negro and Schorfheide, 2011).

Identifying restrictions. Let $\mathbf{y}_t = (FFR_t, GDP_t, GDPDEF_t, COM_t, TR_t, NBR_t)'$. The monetary policy shock is ε_{1t} and the first equation of the SVAR can be interpreted as the monetary policy reaction function. ACR19 set-identify the monetary policy shock using a mixture of sign and zero restrictions on the monetary policy reaction function. The zero restrictions are that FFR_t does not react contemporaneously to TR_t or NBR_t, which implies $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{5,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{5,6})'\mathbf{q}_1 = 0$ and $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{6,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{6,6})'\mathbf{q}_1 = 0$. The matrix containing the coefficients of the zero restrictions is $\mathbf{F}(\boldsymbol{\phi}) = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{5,6}, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{6,6})'$. The sign restrictions are that, all else equal, FFR_t is not decreased in response to higher GDP_t or $GDPDEF_t$, which – given the sign normalization $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{1,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,6})'\mathbf{q}_1 \geq 0$ – implies $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{2,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{2,6})'\mathbf{q}_1 \leq 0$ and $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{3,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{3,6})'\mathbf{q}_1 \leq 0$. The impact response of FFR_t

[10] All results are obtained using MATLAB 2020b on a laptop with Windows 10, an Intel Core i7-6700HQ CPU @ 2.60 GHz with four cores, and 8 GB of RAM.

to the monetary policy shock is also restricted to be nonnegative, which requires $\mathbf{e}_{1,6}'\mathbf{A}_0^{-1}\mathbf{e}_{1,6} = \mathbf{e}_{1,6}'\mathbf{\Sigma}_{tr}\mathbf{q}_1 \geq 0$. The matrix containing the coefficients of the sign restrictions is $\mathbf{S}(\boldsymbol{\phi}) = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{1,6}, -\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{2,6}, -\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{3,6}, (\mathbf{e}_{1,6}'\mathbf{\Sigma}_{tr})')'$.

I also consider other sets of identifying restrictions that add additional sign restrictions to $\mathbf{S}(\boldsymbol{\phi})$. Specifically, I add the sign restrictions on impulse responses proposed in Uhlig (2005). These restrictions are that the impulse response of FFR_t to the monetary policy shock is nonnegative for $h = 0, 1, \ldots, H$ and the impulse responses of $GDPDEF_t$, COM_t and NBR_t are nonpositive for $h = 0, 1, \ldots, H$, where H is a specified horizon. To explore how the algorithms perform under different numbers of sign restrictions, I consider $H \in \{5, 11, 23\}$. In total, there are 27 sign restrictions when $H = 5$, 51 sign restrictions when $H = 11$ and 99 sign restrictions when $H = 23$.

Determining emptiness of $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F}, \mathbf{S})$. Given 1,000 draws from the posterior of $\boldsymbol{\phi}$, I check whether the identified set is empty using Algorithm 3.4.1 and two alternative approaches. I compare the accuracy and speed of the three algorithms.

The first alternative is a rejection-sampling (RS) approach similar to that used by Arias et al. (2018) and Giacomini and Kitagawa (2021) to draw values of $\mathbf{Q}$. The algorithm draws $\mathbf{q}_1$ from a uniform distribution over $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F}) = \left\{\mathbf{q} \in \mathbb{S}^{n-1} : \mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}\right\}$ and checks whether the draw satisfies the sign restrictions.[11] If no draws of $\mathbf{q}_1$ satisfy the sign restrictions after $100,000$ draws, I approximate the identified set as empty. This algorithm may incorrectly classify the identified set as being empty, particularly when draws satisfying the zero restrictions satisfy the sign restrictions with low probability.

The second algorithm is from Giacomini, Kitagawa and Volpicella (2022). Their algorithm relies on the fact that any nonempty identified set for $\mathbf{q}_1$ must contain a vertex on the unit sphere where at least $n-1$ restrictions are binding. The algorithm determines whether the identified set is nonempty by considering all possible combinations of $n-r-1$ binding sign restrictions and checking whether the

[11] The algorithm draws $\mathbf{z} \sim N(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$ and computes $\tilde{\mathbf{q}}_1 = \left[\mathbf{I}_n - \mathbf{F}_1'(\mathbf{F}_1\mathbf{F}_1')^{-1}\mathbf{F}_1\right]\mathbf{z}$, so $\tilde{\mathbf{q}}_1$ satisfies the zero restrictions. $\tilde{\mathbf{q}}_1$ is then normalized so that it satisfies the sign normalization and has unit length before checking whether it satisfies the remaining sign restrictions.

implied vertex satisfies the remaining sign restrictions. This approach will exactly determine whether the identified set is empty, but may become computationally burdensome when the number of sign restrictions is large, since it requires checking $\binom{s}{n-r-1}$ combinations of restrictions before concluding the identified set is empty.

Table 3.1: Determining Emptiness of $\mathcal{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$

	$\Pr(\mathcal{Q}_1(\boldsymbol{\phi}\|\mathbf{F},\mathbf{S}) = \emptyset)$ (%)			Computing Time (s)		
Restrictions	A4.1	RS	GKV	A4.1	RS	GKV
(1)	0.00	0.00	0.00	9.86	0.04	0.17
(2)	0.60	1.00	0.60	9.26	6.55	4.26
(3)	6.50	8.00	6.50	9.38	50.55	103.78
(4)	31.60	35.40	31.60	9.64	223.51	3,277.2

Note: (1) are the restrictions from ACR19 (2 zero restrictions, 4 sign restrictions); (2), (3) and (4) are the restrictions from ACR19 plus the restrictions from Uhlig (2005) with $H = 5$ (27 sign restrictions), $H = 11$ (51 sign restrictions) and $H = 23$ (99 sign restrictions), respectively; A4.1 refers to Algorithm 4.1; RS refers to the rejection-sampling approach; GKV refers to the algorithm from Giacomini, Kitagawa and Volpicella (2022).

To compare the accuracy of the algorithms, I compute the posterior probability that the identified set is empty. To compare computational efficiency, I tabulate the time taken to check whether the identified set is empty. Under the restrictions from ACR19, all three algorithms correctly determine that the identified set is nonempty at every draw of $\boldsymbol{\phi}$ (Table 3.1). Although Algorithm 3.4.1 is slower than the two alternatives in this case, in practice it would not be necessary to numerically check whether the identified set is nonempty, because the identified set is never empty when $r+s \leq n$.[12] As the number of restrictions increases, the rejection-sampling approach misclassifies the identified set as being empty at some draws of $\boldsymbol{\phi}$. Algorithm 3.4.1 is somewhat slower than the two alternatives under the second set of restrictions, but is much faster under the two larger sets of restrictions.

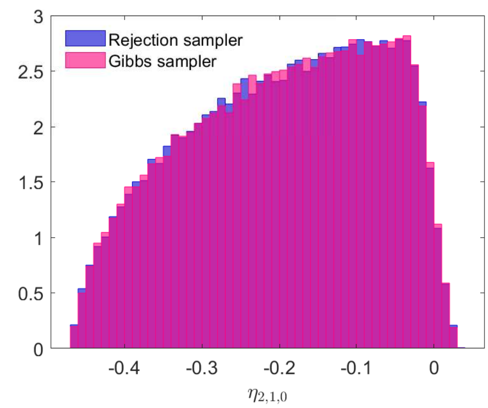
Obtaining uniform draws over $\mathcal{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$. I verify that the Gibbs sampler generates draws from the uniform distribution over $\mathcal{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ by comparing the distributions of draws obtained using the Gibbs sampler and the rejection sampler.

[12]This can be shown by applying Gordan's Theorem (e.g. p. 31 of Mangasarian (1969)) after transforming the system of identifying restrictions into a system of sign restrictions in a lower-dimensional space.

Experiments using parallel Markov chains initialized at the same value suggest that dropping the first three draws from the Gibbs sampler is sufficient to remove dependence on the initial value; the distribution of the fourth draw across the parallel chains is statistically indistinguishable from the distribution generated by the rejection sampler (which generates independent draws). Dropping every second draw from the Gibbs sampler is sufficient to eliminate a significant first-order autocorrelation in the original set of draws.

Under the restrictions from ACR19 and given a single (random) draw of $\boldsymbol{\phi}$ with nonempty identified set, I obtain 100,000 draws of $\mathbf{q}$ from the Gibbs sampler after dropping the initial three draws and keeping every second draw. Figure 3.1 plots histograms of the impact impulse response of output obtained using the two samplers; the distributions appear very similar and a two-sample Kolmogorov-Smirnov test fails to reject the null hypothesis that the two sets of draws are generated by the same distribution (p-value $= 0.8$).

Figure 3.1: Histogram of Impulse Response Under Alternative Sampling Algorithms



Note: Impact response of output; based on 100,000 draws of $\mathbf{q}_1$ at random draw of $\boldsymbol{\phi}$.

Next, I compare the efficiency of the two algorithms. I embed the Gibbs sampler (with a burn-in of three draws) and the rejection sampler within a standard posterior sampler for $\boldsymbol{\phi}$ to obtain 1,000 draws from the joint posterior of $(\boldsymbol{\phi}, \mathbf{q}_1)$ such that $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F}, \mathbf{S})$ is nonempty. The rejection sampler is more efficient than the Gibbs sampler when there are few sign restrictions (Table 3.2). The algorithms perform similarly when there is an intermediate number of sign restrictions. Under the larger sets of restrictions, the Gibbs sampler is more efficient.

Table 3.2: Time Taken to Obtain 1,000 Draws (s)

	Draws of $(\boldsymbol{\phi}, \mathbf{q}_1)$		Draws of $IS_\eta(\boldsymbol{\phi}\|F,S)$		
Restrictions	Gibbs	RS	A4.1	RS	GMM18
(1)	13	3	407	396	10
(2)	13	13	591	609	1537
(3)	14	61	759	797	10,638
(4)	19	352	887	1,270	85,990

Note: (1) are the restrictions from ACR19 (2 zero restrictions, 4 sign restrictions); (2), (3) and (4) are the restrictions from ACR19 plus the restrictions from Uhlig (2005) with $H = 5$ (27 sign restrictions), $H = 11$ (51 sign restrictions) and $H = 23$ (99 sign restrictions), respectively; 'Gibbs' refers to Algorithm 4.2 with a burn-in of three draws; RS refers to the rejection-sampling approach; GMM18 refers to the active-set algorithm from Gafarov et al. (2018).

Computing the bounds of $IS_\eta(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$. At 1,000 draws of $\boldsymbol{\phi}$ where $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ is nonempty, I compute the lower and upper bounds of $IS_\eta(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ when $\eta = \eta_{i,j,h}(\boldsymbol{\phi},\mathbf{q}_1)$ is the output response to the monetary policy shock at horizons $h = 0,\ldots,60$. The upper bound, $u(\boldsymbol{\phi})$, is defined as the value function of the optimization problem $u(\boldsymbol{\phi}) = \max_{\mathbf{q}_1 \in \mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})} \eta_{i,j,h}(\boldsymbol{\phi},\mathbf{q}_1)$, which is a quadratically constrained linear program with linear equality and inequality constraints. The lower bound, $l(\boldsymbol{\phi})$, is defined as the value function from the corresponding minimization problem.

I consider three alternative approaches. The first uses Algorithm 4.1 to check whether the identified set is nonempty and to obtain a value of $\mathbf{q}_1$ satisfying the identifying restrictions, which is used to initialize a gradient-based numerical optimization routine.[13] The second algorithm uses the same numerical optimization routine, but uses the rejection-sampling approach to obtain the initial value of $\mathbf{q}_1$. The third approach uses the active-set algorithm described in Gafarov et al. (2018).[14] This ap-

[13] I use the interior-point algorithm in MATLAB's fmincon optimizer with analytical gradients of the objective function and constraints.

[14] Given a set of binding restrictions, Gafarov et al. (2018) derive an expression for the value function (up to sign) of the optimization problems that define the bounds of the identified set. They also provide expressions for the corresponding solutions of the problems. They propose computing the value function at every possible combination of binding restrictions and checking whether the corresponding solution satisfies the remaining sign restrictions. The lower and upper bounds of the identified set are then the minimum and maximum, respectively, over the feasible value functions obtained at each combination of binding restrictions.

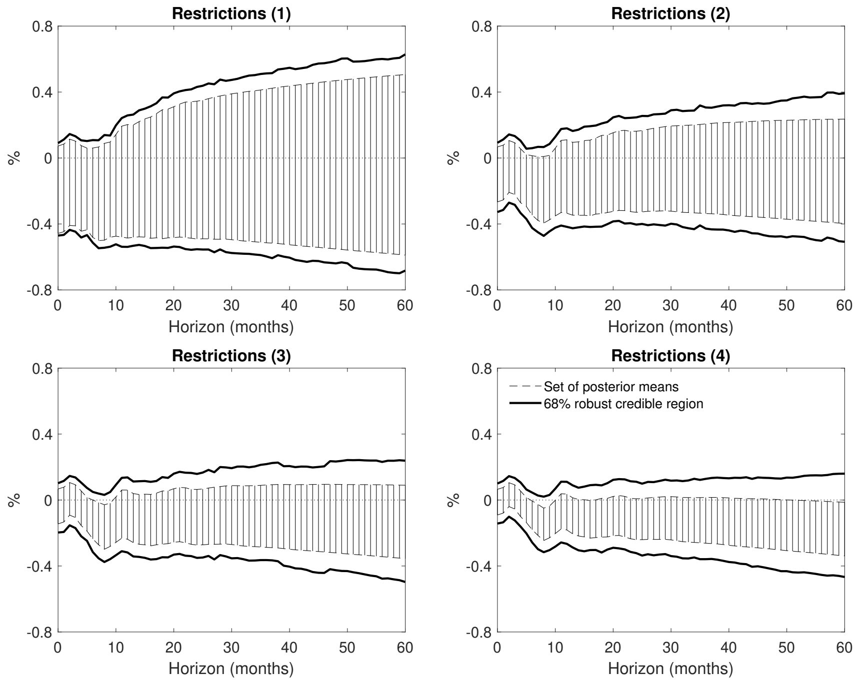
proach may be computationally burdensome when the number of sign restrictions is large, since it requires computing the bounds of the identified set at $\sum_{k=0}^{n-r-1}\binom{s}{k}$ combinations of binding sign restrictions.

When there is a small number of sign restrictions, the algorithm from Gafarov et al. (2018) is the most efficient of the three (Table 3.2). The other two algorithms perform similarly, since the bulk of the computing time is spent on the optimization step – which is common across the two approaches – rather than on trying to find a feasible initial value. As the number of sign restrictions increases, the algorithm from Gafarov et al. (2018) becomes computationally burdensome due to the explosion in the number of combinations of active restrictions to check. When $H = 23$, using Algorithm 4.1 to obtain a feasible initial value for the numerical optimization routine is about 30 per cent faster than obtaining the initial value via rejection sampling.

To provide some sense of how tight the identified set is on average under the different identifying restrictions, Figure 3.2 plots the set of posterior means and 68 per cent robust credible regions for the output response. These quantities are proposed by Giacomini and Kitagawa (2021) to assess or eliminate the sensitivity of posterior inference in set-identified models to the choice of conditional prior for $\mathbf{Q}|\boldsymbol{\phi}$. The set of posterior means is the average of $IS_{\eta}(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ over the posterior for $\boldsymbol{\phi}$ and can be interpreted as a consistent estimator of the identified set. The robust credible region is the shortest interval covering 68 per cent of the posterior distribution under all possible conditional priors for $\mathbf{Q}|\boldsymbol{\phi}$ that satisfy the identifying restrictions, and can be interpreted as an asymptotically valid frequentist confidence interval. Each additional set of sign restrictions appears to appreciably truncate the identified set, on average. This explains the improvement in the performance of the proposed algorithms relative to those based on rejection sampling as the number of restrictions increases.

An extremely large number of restrictions. This section provides an example of a set of identifying restrictions under which (accurate) posterior inference would be extremely computationally burdensome using existing algorithms. Specifically, in

Figure 3.2: Output Response to a Monetary Policy Shock



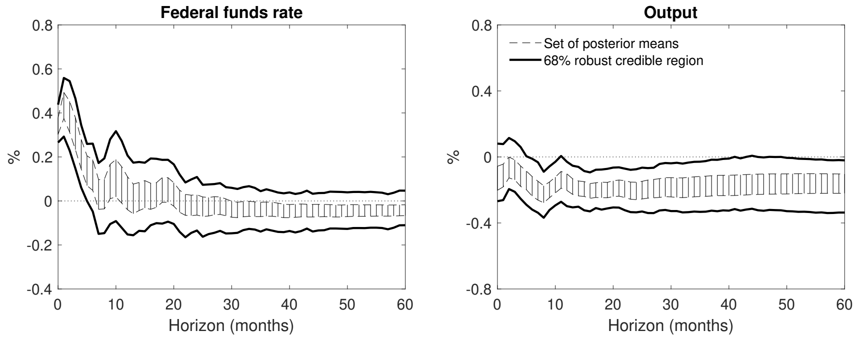
Note: (1) are the restrictions from ACR19 (2 zero restrictions, 4 sign restrictions); (2), (3) and (4) are the restrictions from ACR19 plus the restrictions from Uhlig (2005) with $H = 5$ (27 sign restrictions), $H = 11$ (51 sign restrictions) and $H = 23$ (99 sign restrictions), respectively; responses are to a positive standard-deviation shock to the federal funds rate and are obtained using Algorithm 4.1 and a gradient-based numerical optimization routine.

addition to the restrictions in ACR19 and the restrictions in Uhlig (2005) when $H = 5$, I impose a narrative restriction on the timing of the maximum realisation of the monetary policy shock, as in Giacomini et al. (2021). This 'shock-rank' restriction is that the monetary policy shock in October 1979 – the month in which Paul Volcker dramatically and unexpectedly raised the federal funds rate – was the largest positive realization of the monetary policy shock in the sample. This restriction requires that $\varepsilon_{1k} = \mathbf{e}_{1,n}'\mathbf{A}_0\mathbf{u}_k = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{u}_k)'\mathbf{q}_1 \geq 0$ and $\varepsilon_{1k} \geq \max_{t \neq k}\{\varepsilon_{1t}\}$ where k is the index corresponding to October 1979. The latter restriction is equivalent to $(\mathbf{\Sigma}_{tr}^{-1}(\mathbf{u}_k - \mathbf{u}_t))'\mathbf{q}_1 \geq 0$ for $t \neq k$. The restriction generates T additional inequality restrictions on $\mathbf{q}_1$, where T is the sample size and the inequality restrictions depend on the data (via the reduced-form VAR innovations).

The restriction noticeably tightens the set of posterior means and robust credible intervals for the output response; the set of posterior means excludes zero at all horizons considered and the robust credible intervals exclude zero at most horizons

(Figure 3.3). The posterior lower probability – the smallest probability attainable in the class of posteriors – of a negative output response at the two-year horizon is around 95 per cent. The restriction therefore appears to be extremely informative when combined with the restrictions from ACR19 and Uhlig (2005). However, the posterior probability that the identified set is empty is very high (around 96 per cent), which suggests that the restriction is inconsistent with the data.

Figure 3.3: Impulse Responses to Monetary Policy Shock – Shock-rank Restriction



Note: Shock-rank restriction is the restriction that the monetary policy shock in October 1979 was the largest positive realization of the shock in the sample; this restriction is in addition to the restrictions from ACR and Uhlig (2005) with $H = 5$ (525 sign restrictions in total); responses are to a positive standard-deviation shock to the federal funds rate and are obtained using Algorithm 4.1 and a numerical optimization procedure.

The large system of inequality restrictions and narrow identified set generated by the restriction poses difficulties for existing algorithms. The approach in GKV would require checking $\binom{498+3+24}{3} = 23,979,550$ combinations of sign restrictions to determine that the identified set is empty. The approach in GMM would require considering $\sum_{k=0}^{3}\binom{498+3+24}{k} = 24,117,626$ combinations of restrictions for every parameter of interest (i.e, for every variable and every impulse-response horizon) to compute the bounds of the identified set. Furthermore, given how tight the identified set appears to be on average over the posterior for $\boldsymbol{\phi}$, the rejection-sampling approach would require a very large number of draws of $\mathbf{q}_1$ given the zero restrictions to accurately approximate the posterior probability that the identified set is empty or to conduct inference under a conditionally uniform prior.

Discussion. Overall, the results of this empirical exercise suggest that the new algorithms are likely to be preferable to existing alternatives when more than a handful of sign restrictions are imposed and/or these sign restrictions substantially truncate the identified set given the zero restrictions. The new algorithms should therefore facilitate the use of rich sets of sign restrictions alongside zero restrictions. It is difficult to provide definitive guidance about the number of restrictions above which the new algorithms will be more computationally efficient than the alternative algorithms considered; for example, whether the new algorithms are more efficient than the alternatives based on rejection sampling will depend on the extent to which the sign restrictions truncate the identified set given the zero restrictions. If practitioners want to use the fastest algorithm in any particular circumstance, they could test the algorithms against each other at particular values of the reduced-form parameters, such as at the maximum-likelihood estimate or at a small number of draws from the posterior.

3.6 Extensions

The algorithms described in the paper can be extended to some additional cases where there are restrictions on multiple columns of $\mathbf{Q}$.

3.6.1 Some columns of Q are point-identified

Consider the case where the first i^* columns of $\mathbf{Q}$, $[\mathbf{q}_1, \ldots, \mathbf{q}_{i^*}]$, are point-identified by zero restrictions, but interest is in impulse responses to the $j^* = (i^*+1)$th shock with both zero and sign restrictions on $\mathbf{q}_{j^*}$. Let $\mathbf{F}^{(i)}(\boldsymbol{\phi})$ represent the $r_i \times n$ matrix containing the coefficients of the zero restrictions constraining $\mathbf{q}_i$ and let

$$\mathbf{F}(\boldsymbol{\phi}, \mathbf{Q}) = \begin{bmatrix} \mathbf{F}^{(1)}(\boldsymbol{\phi})\mathbf{q}_1 \\ \vdots \\ \mathbf{F}^{(n)}(\boldsymbol{\phi})\mathbf{q}_n \end{bmatrix} = \mathbf{0}_{\sum_{i=1}^{n} r_i \times 1}. \tag{3.10}$$

If the zero restrictions do not constrain $\mathbf{q}_i$, then $\mathbf{F}^{(i)}(\boldsymbol{\phi})$ does not exist and $r_i = 0$. As in Giacomini and Kitagawa (2021), assume the variables in $\mathbf{y}_t$ are ordered to satisfy the following ordering convention.

Definition 3.6.1. Ordering convention. Order the variables in $\mathbf{y}_t$ so that r_i satisfies $r_1 \geq r_2 \geq \ldots \geq r_n \geq 0$. If the impulse response of interest is to the j^*th variable, order the j^*th variable first among ties.

A necessary condition for exact identification of the first i^* columns of $\mathbf{Q}$ is that $\text{rank}(\mathbf{F}^{(i)}(\boldsymbol{\phi})) = r_i = n - i$ for $i = 1, \ldots, i^*$ (Rubio-Ramírez et al. (2010)). A necessary and sufficient condition is that

$$\text{rank}\left(\left[\mathbf{F}^{(i)}(\boldsymbol{\phi})', \mathbf{q}_1, \ldots, \mathbf{q}_{i-1}\right]'\right) = n - 1, \quad \text{for } i = 1, \ldots, i^*, \tag{3.11}$$

which additionally requires that the restrictions in $\mathbf{F}^{(i)}(\boldsymbol{\phi})$, $i = 1, \ldots, i^*$, are 'non-redundant' in the sense discussed in Bacchiocchi and Kitagawa (2021).[15]

Assume $r_{j^*} < n - j^*$, in which case $\mathbf{q}_{j^*}$ is set-identified. $\mathbf{q}_{j^*}$ is constrained by the sign restrictions $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_{j^*} \geq \mathbf{0}_{s\times 1}$ and an extended set of zero restrictions incorporating the restriction that $\mathbf{q}_{j^*}$ is orthogonal to the preceding columns of $\mathbf{Q}$:

$$\ddot{\mathbf{F}}^{(j^*)}(\boldsymbol{\phi})\mathbf{q}_{j^*} = \left[\mathbf{F}^{(j^*)}(\boldsymbol{\phi})', \mathbf{q}_1, \ldots, \mathbf{q}_{i^*}\right]' \mathbf{q}_{j^*} = \mathbf{0}_{(r_{j^*}+j^*-1)\times 1}, \tag{3.12}$$

where I have suppressed the dependence of $\ddot{\mathbf{F}}^{(j^*)}(\boldsymbol{\phi})$ on $\mathbf{F}^{(i)}(\boldsymbol{\phi})$, $i = 1, \ldots, i^*$. If the conditions for exact identification above are satisfied, each $\mathbf{q}_i$, $i = 1, \ldots, i^*$, can be determined iteratively as follows. First, find a unit-length vector $\mathbf{q}_1$ satisfying $\mathbf{F}^{(1)}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{(n-1)\times 1}$ by computing an orthonormal basis for the null space of $\mathbf{F}^{(1)}(\boldsymbol{\phi})$. Normalize $\mathbf{q}_1$ so it satisfies the sign normalization $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\mathbf{q}_1 \geq 0$. Then, for $i = 2, \ldots, i^*$, find a unit-length vector $\mathbf{q}_i$ satisfying $(\mathbf{F}^{(i)}(\boldsymbol{\phi})', \mathbf{q}_1, \ldots, \mathbf{q}_{i-1})'\mathbf{q}_i = \mathbf{0}_{(n-1)\times 1}$ by computing an orthonormal basis for the null space of $(\mathbf{F}^{(i)}(\boldsymbol{\phi})', \mathbf{q}_1, \ldots, \mathbf{q}_{i-1})'$. Normalize $\mathbf{q}_i$ so it satisfies the sign normalization $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{i,n})'\mathbf{q}_i \geq 0$. Under the assumption that $\text{rank}(\ddot{\mathbf{F}}^{(j^*)}(\boldsymbol{\phi})) = r_{j^*} + j^* - 1$, $\ddot{\mathbf{F}}^{(j^*)}(\boldsymbol{\phi})$ can replace $\mathbf{F}(\boldsymbol{\phi})$ in the algorithms described in Section 3.4 without further modification.

[15] If the condition in Equation (3.11) is satisfied, then $\mathbf{q}_i$ is restricted to lie in the one-dimensional linear subspace of $\mathbb{R}^n$ satisfying $(\mathbf{F}^{(i)}(\boldsymbol{\phi})', \mathbf{q}_1, \ldots, \mathbf{q}_{i-1})'\mathbf{q}_i = \mathbf{0}_{(n-1)\times 1}$. The sign normalization and requirement that $\mathbf{q}_i$ lie on the unit sphere pin down $\mathbf{q}_i$.

3.6.2 A subset of the columns of Q is determined up to a linear subspace

The algorithms can also be applied when $(\mathbf{q}_1, \dots, \mathbf{q}_{i^*})$ is not point-identified, but is pinned down to lie within an i^*-dimensional linear subspace of $\mathbb{R}^n$. Consider the case where $\mathbf{F}^{(1)}(\boldsymbol{\phi}) = \dots = \mathbf{F}^{(i^*)}(\boldsymbol{\phi})$, so $r_1 = \dots = r_{i^*} \equiv \tilde{r}$, and assume $\tilde{r} = n - i^*$, which implies $\mathbf{q}_i$ is set-identified for $i = 1, \dots, i^*$. For example, this pattern of restrictions arises in proxy SVARs when there are multiple proxies for multiple shocks (e.g. Giacomini et al. (2022a)). The restriction $\mathbf{F}^{(i)}(\boldsymbol{\phi})\mathbf{q}_i = \mathbf{0}_{\tilde{r}\times 1}$ restricts $\mathbf{q}_i$ to lie in a linear subspace of $\mathbb{R}^n$ with dimension $n - \tilde{r} = i^*$. Since $\mathbf{F}^{(i)}(\boldsymbol{\phi})$ is common for $i = 1, \dots, i^*$, the first i^* columns of $\mathbf{Q}$ are restricted to lie in the same i^*-dimensional subspace. An orthonormal basis for this subspace is $N(\mathbf{F}^{(1)}(\boldsymbol{\phi}))$. $\mathbf{q}_j^*$ must be orthogonal to the preceding columns of $\mathbf{Q}$, so it must be orthogonal to the i^*-dimensional subspace spanned by the columns of $N(\mathbf{F}^{(1)}(\boldsymbol{\phi}))$. $\mathbf{q}_{j^*}$ must therefore satisfy $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_{j^*} \geq \mathbf{0}_{s\times 1}$ and the extended set of zero restrictions

$$\ddot{\mathbf{F}}^{(j^*)}(\boldsymbol{\phi})\mathbf{q}_{j^*} = \begin{bmatrix} \mathbf{F}^{(j^*)}(\boldsymbol{\phi}) \\ N(\mathbf{F}^{(1)}(\boldsymbol{\phi}))' \end{bmatrix} \mathbf{q}_{j^*} = \mathbf{0}_{(r_{j^*}+j^*-1)\times 1}. \tag{3.13}$$

Under the assumption $\text{rank}(\ddot{\mathbf{F}}^{(j^*)}(\boldsymbol{\phi})) = r_{j^*} + j^* - 1 < n - j^*$, $\ddot{\mathbf{F}}^{(j^*)}(\boldsymbol{\phi})$ can replace $\mathbf{F}(\boldsymbol{\phi})$ in the algorithms described in Section 3.4 without further modification.

3.6.3 Sign and zero restrictions on multiple columns of Q

Consider the case where there are are zero and sign restrictions on the first $i^* < n$ columns of $\mathbf{Q}$ with $0 \leq r_i < n - i$ for $i = 1, \dots, i^*$, so $\mathbf{q}_i$ is set-identified for all $i = 1, \dots, n$. Let $\mathbf{S}^{(i)}(\boldsymbol{\phi})\mathbf{q}_i \geq \mathbf{0}_{s_i\times 1}$ represent the sign restrictions constraining $\mathbf{q}_i$ and let $\mathbf{S}(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{\sum_{i=1}^n s_i \times 1}$ represent the system of sign restrictions (including the sign normalizations). The identified set for $\mathbf{Q}$ is

$$\mathscr{Q}(\boldsymbol{\phi}|\mathbf{F}, \mathbf{S}) = \left\{ \mathbf{Q} \in \mathscr{O}(n) : \mathbf{F}(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{\sum_{i=1}^n r_i \times 1}, \mathbf{S}(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{\sum_{i=1}^n s_i \times 1} \right\}. \tag{3.14}$$

When there are sign restrictions only, AD21 provide a sufficient condition for checking whether $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ is empty and a sufficient condition for checking whether it is nonempty. These sufficient conditions can be extended to the case with zero restrictions in the following way.

First, a sufficient condition for $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{F},\mathbf{S}) = \emptyset$ is $\mathscr{Q}_i(\boldsymbol{\phi}|\mathbf{F}^{(i)},\mathbf{S}^{(i)}) = \emptyset$ for any $i \in \{1,\ldots,i^*\}$. Intuitively, if the identified set for a single column of $\mathbf{Q}$ is empty when imposing only the restrictions directly constraining that column, the identified set for $\mathbf{Q}$ itself must be empty. To check this, one can apply Algorithm 3.4.1 for each i, replacing $\mathbf{F}(\boldsymbol{\phi})$ and $\mathbf{S}(\boldsymbol{\phi})$ with $\mathbf{F}^{(i)}(\boldsymbol{\phi})$ and $\mathbf{S}^{(i)}(\boldsymbol{\phi})$, respectively.

Second, a sufficient condition for $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$ to be nonempty is that $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F}^{(1)},\mathbf{S}^{(1)})$ is nonempty and $\mathscr{Q}_i(\boldsymbol{\phi}|\hat{\mathbf{F}}^{(i)},\mathbf{S}^{(i)})$ is nonempty for all $i = 2,\ldots,i^*$, where

$$\hat{\mathbf{F}}^{(i)}(\boldsymbol{\phi}) = \left[\mathbf{F}^{(i)}(\boldsymbol{\phi})', \mathbf{q}_1^{(0)}, \ldots, \hat{\mathbf{q}}_{i-1}^{(0)}\right]'. \tag{3.15}$$

$\mathbf{q}_1^{(0)}$ is the transformed Chebyshev centre obtained from applying Algorithm 3.4.1 under the system of restrictions $\mathbf{F}^{(1)}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r_1\times 1}$ and $\mathbf{S}^{(1)}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s_1\times 1}$, and $\hat{\mathbf{q}}_i^{(0)}$ is the transformed Chebyshev centre obtained from applying Algorithm 3.4.1 under the system of restrictions $\hat{\mathbf{F}}^{(i)}(\boldsymbol{\phi})\mathbf{q}_i = \mathbf{0}_{r_i\times 1}$ and $\mathbf{S}^{(i)}(\boldsymbol{\phi})\mathbf{q}_i \geq \mathbf{0}_{s_i\times 1}$. It is only necessary to check this condition for the first i^* columns of $\mathbf{Q}$, since a set of $n - i^*$ orthonormal vectors satisfying the sign normalizations can always be constructed in the null space of $(\mathbf{q}_1^{(0)}, \ldots, \mathbf{q}_{i^*}^{(0)})$.

If neither sufficient condition is satisfied, one could attempt to determine whether the identified set is nonempty using rejection sampling. However, as in the case where a single column of $\mathbf{Q}$ is restricted, this is likely to be inaccurate or computationally burdensome when the sign restrictions markedly tighten the identified set for $\mathbf{Q}$ given the zero restrictions.

AD21 also describe a Gibbs sampler that is applicable when there are sign restrictions on multiple columns of $\mathbf{Q}$. Similar to the case where a single column of $\mathbf{Q}$ is restricted, this sampler can be extended to allow for zero restrictions. Assume one is able to obtain a value of $\mathbf{Q}_{1:i^*} = (\mathbf{q}_1, \ldots, \mathbf{q}_{i^*})$ satisfying the sign and zero restrictions and that the parameter of interest is a function of $\mathbf{Q}_{1:i^*}$ (e.g. an impulse response to one of the first i^* shocks). Also, assume that $r_i < n - i^*$ for all $i = 1, \ldots, n$ and that $i^* < n - 1$.

Algorithm 3.6.1. Gibbs sampler for $\mathbf{Q}_{1:i^*}$. Assume $\mathbf{Q}_{1:i^*}^{(0)} = (\mathbf{q}_1^{(0)}, \ldots, \mathbf{q}_{i^*}^{(0)})$ is available satisfying the system of identifying restrictions. Let L be the desired number of draws of $\mathbf{Q}_{1:i^*}$. For each $k = 1, \ldots, L$, sequentially complete the following steps for $j = 1, \ldots, i^*$:

- ***Step 1.*** Compute $\mathbf{F}^\dagger = (\mathbf{q}_1^{(k)}, \ldots, \mathbf{q}_{j-1}^{(k)}, \mathbf{q}_{j+1}^{(k-1)}, \ldots, \mathbf{q}_{i^*}^{(k-1)}, \mathbf{F}_j')'$.
- Compute the change-of-basis matrix $\mathbf{K} = (N(\mathbf{F}^\dagger), N(N(\mathbf{F}^\dagger)'))$ and transform the coefficient vectors of the sign, zero and orthogonality restrictions into the new basis via $\tilde{\mathbf{S}}_j = (\mathbf{K}^{-1}\mathbf{S}_j')'$ and $\tilde{\mathbf{F}} = (\mathbf{K}^{-1}\mathbf{F}^{\dagger\prime})'$.
- ***Step 2.*** Project the coefficient vectors of the sign restrictions in the new basis onto the linear subspace spanned by the rows of $\tilde{\mathbf{F}}$ and drop the last $r_j + i^* - 1$ elements of the resulting vectors. The transformed matrix of coefficients is $\bar{\mathbf{S}}_j = (\mathbf{M}(\mathbf{I}_n - \tilde{\mathbf{F}}'(\tilde{\mathbf{F}}\tilde{\mathbf{F}}')^{-1}\tilde{\mathbf{F}})\tilde{\mathbf{S}}')'$, where $\mathbf{M} = (\mathbf{I}_{n-r_j-i^*-1}, \mathbf{0}_{(n-r_j-i^*-1)\times(r_j+i^*+1)})$.
- ***Step3.*** Set $\mathbf{z}_j^{(k-1)} = \mathbf{M}\mathbf{K}^{-1}\mathbf{q}_j^{(k-1)}$ and apply Steps 1 and 2 of Algorithm 4.2 for $i = 1, \ldots, n - r_j - i^* + 1$ to obtain $\mathbf{q}_j^{(k)}$.

The key difference between this algorithm and Algorithm 4.2 is that, within every iteration of the Gibbs sampler, the columns of $\mathbf{Q}_{1:i^*}$ other than $\mathbf{q}_j$ are treated as given. The condition that $\mathbf{q}_j$ is orthogonal to the remaining columns of $\mathbf{Q}_{1:i^*}$ can be treated as a set of zero restrictions. An important assumption is that an initial value of $\mathbf{Q}_{1:i^*}$ satisfying the identifying restrictions is available; when the sufficient condition for a nonempty identified set described above is not satisfied, obtaining such a value may require the use of rejection-sampling methods. Numerical exercises indicate that this algorithm draws from the same uniform distribution over the identified set for $\mathbf{Q}_{1:i^*}$ as the rejection samplers described in Arias et al. (2018) and Giacomini and Kitagawa (2021). The algorithm is likely to be more efficient than rejection sampling when sign restrictions considerably truncate the identified set for

$\mathbf{Q}_{1:i^*}$ given the zero restrictions. The algorithm may be useful for approximating the bounds of the identified set for a scalar parameter of interest.[16]

3.7 Conclusion

In SVAR models, a system of sign and zero restrictions constraining a single column of the orthonormal matrix can be expressed as a system of sign restrictions in a lower-dimensional space. Consequently, algorithms that are useful for conducting Bayesian inference under sign restrictions can be extended to the case where there are also zero restrictions. I show that such algorithms can be more accurate and computationally efficient than existing alternatives, particularly when a large number of sign restrictions considerably truncates the identified set given the zero restrictions. The algorithms in this paper should therefore facilitate Bayesian inference when rich sets of sign restrictions are imposed alongside zero restrictions.

[16]When multiple columns of $\mathbf{Q}$ are constrained by the identifying restrictions, the approach from Gafarov et al. (2018) to computing the bounds of the identified set is generally inapplicable. Moreover, computing these bounds via numerical optimization may be difficult, since the optimization problem is nonconvex. For a rejection-sampling approach to this problem, see Algorithm 2 of Giacomini and Kitagawa (2021).

Chapter 4

Robust Bayesian Analysis for Econometrics

4.1 Introduction

Bayesian analysis has many attractive features, such as finite-sample decision-theoretic optimality and computational tractability. The crucial assumption to enjoy these benefits is that the researcher can specify the inputs of the analysis: the likelihood, a prior distribution for the parameters, and a loss function if the analysis involves statistical decisions. In practice, however, researchers commonly face uncertainty about the choice of these inputs. Robust Bayesian analysis addresses this uncertainty by quantifying the sensitivity of the results of Bayesian inference to changes in these inputs. In this paper we present a selective review of the literature on robust Bayesian analysis.[1]

How sensitivity should be measured in a robust Bayesian analysis depends on how the input is perturbed. A local approach quantifies marginal changes in posterior quantities with respect to local perturbation in the input; for example, see Gustafson (2000), Müller (2012), and the references therein. In contrast, a global approach introduces a set of inputs and summarizes posterior sensitivity by reporting the corresponding set of posterior quantities.

The main focus of this paper is on sensitivity to the prior input and on global,

[1] See Berger (1994) and Ríos Insua and Ruggeri (2000) for previous surveys on the topic.

rather than local, robust Bayesian analysis. There are several reasons for pursuing a global approach. First, the set of prior distributions to be specified as an input of the robust analysis can be viewed as representing ambiguous beliefs (Knightian uncertainty), which has been well studied in economic decision theory and experimental economics since the pioneering works of Ellsberg (1961) and Schmeidler (1989). Some statisticians have also argued that a set of priors is easier to elicit than a single prior (Good (1965)). Second, unless approximating the posterior mean with a particular local perturbation of the prior is of the main interest, sets of posterior means or probabilities are easier to interpret than local sensitivity parameters such as the derivative of the posterior mean with respect to the prior mean of a parameter (e.g., the set of plausible posteriors may be too large for low-dimensional local approximations to perform well). Third, although it is often argued that the set of posteriors is more difficult to compute than local sensitivity parameters, this is not the case for the structural vector autoregression (SVAR) models considered in detail in this paper.

We first consider a general environment for inference and statistical decision-making under multiple priors and discuss different ways of constructing the set of priors. We then specialize the discussion to set-identified structural models, where the posterior sensitivity is due to a component of the prior (the conditional prior for the structural parameter given the reduced-form parameter) that is never updated by the data. We illustrate the "full-ambiguity" approach of Giacomini and Kitagawa (2021; henceforth, GK) to constructing a set of priors in general set-identified structural models. In addition, we provide new theoretical results that can be used to derive and compute the set of posterior moments for sensitivity analysis and that are also useful for computing the optimal statistical decision in the presence of multiple priors. These results are new to the literature and generalize some results in GK.

The paper ends with a detailed and self-contained discussion of robust Bayesian inference for set-identified SVARs. Set-identification arises in SVARs when there are sign restrictions and/or under-identifying zero restrictions on functions of the structural parameters, or in SVARs identified using external instruments ("proxy SVARs") when there are multiple instruments for multiple shocks (see Giacomini, Kitagawa and Read (2022)). We review three robust Bayesian approaches to inference that are applicable in set-identified SVARs. The common feature of the approaches is that they replace the unrevisable component of the prior with multiple priors. We cast the approaches within a common framework, discuss their numerical implementation and illustrate their use in an empirical example. Our goal is to elucidate how and when these different approaches may be useful in eliminating or quantifying the influence of the unrevisable component of the prior on posterior inference. Ultimately, we argue that the robust Bayesian outputs generated by these methods should be reported alongside the standard Bayesian outputs that are typically reported in studies using set-identified SVARs.

The first approach to robust Bayesian inference for set-identified SVARs is GK, which replaces the unrevisable component of the prior with the set of all priors that are consistent with the imposed identifying restrictions. This generates a set of posteriors, which can be summarised by a set of posterior means and a robust credible region, which is the shortest interval assigned at least a given posterior probability under all posteriors within the set. One can also report the lower or upper posterior probability of some event (e.g, the output response to a monetary policy shock is negative at some horizon), which is the smallest or largest probability of the event over all posteriors in the set. GK show that, under certain conditions, the set of posterior means is a consistent estimator of the identified set and the robust credible region attains valid frequentist coverage of the true identified set asymptotically. In contrast, under standard Bayesian inference, the posterior mean asymptotically lies at a point within the identified set that is determined entirely by the prior, and standard credible intervals for a parameter of interest lie strictly within the identified set asymptotically (Moon and Schorfheide 2012). The approach of GK therefore reconciles the asymptotic disagreement between frequentist and Bayesian inference in set-identified models.

The second approach is the "model-averaging" approach of Giacomini, Kitagawa and Volpicella (2022; henceforth, GKV). The approach extends Bayesian model averaging to a mix of single-prior and multiple-prior models. Given prior probabilities chosen by the user, the multiple-posterior models are averaged (posterior-by-posterior) with the single-posterior models, where the weights on each model are the posterior model probabilities. This averaging generates a set of posteriors, which can be summarised as in GK. For instance, when there is one point-identified model (which yields a single prior) and one set-identified model, the post-averaging set of posterior means shrinks the bounds of the set of posterior means in the set-identified model towards the posterior mean in the point-identified model. GKV explain the conditions under which prior model probabilities are revised, in which case the data may be informative about which identifying restrictions are more plausible.

The third approach is the "KL-neighborhood" approach proposed by Giacomini, Kitagawa and Uhlig (2019; henceforth, GKU). GKU consider a set of priors in a Kullback-Leibler (KL) neighborhood of a 'benchmark' prior. The motivation for this proposal is that one may not want to entertain priors that are, in some sense, far from the benchmark prior, because the benchmark prior may be partially credible. Similarly to GK, this generates a set of posteriors, which can be summarised by a set of posterior means and/or quantiles. GKU also derive a point estimator solving a Bayesian statistical decision problem allowing for ambiguity over the set of priors within the KL-neighborhood around the benchmark prior (the 'posterior Gamma minimax problem').

The robustness issue reviewed in this article focuses exclusively on misspecification of and sensitivity to the prior distribution in the Bayesian setting. There is a vast literature on robust statistics from the frequentist perspective; for example, see Huber and Ronchetti (2009), Rieder (1994), and references therein for classical approaches to robust statistical methods. The frequentist approach to robustness typically concerns misspecifying the likelihood (contamination of the data-generating process), identifying assumptions, moment conditions, or the distribution of unobservables. The main focuses of this literature are to quantify sensitivity of estimation and inference to such mispecification and develop estimators that are robust against it. For recent advances in econometrics, see Kitamura et al. (2013), Andrews et al. (2017, 2020), Bonhomme and Weidner (2018), Christensen and Connault (2021) and Armstrong and Kolesár (2021).

The remainder of the paper is structured as follows. Section 4.2 presents a general overview of robust Bayesian analysis. Section 4.3 specializes the discussion to set-identified structural models. Section 4.4 presents new theoretical results regarding the set of posterior moments. Section 4.5 discusses three approaches to robust Bayesian analysis in set-identified SVARs. Section 4.6 contains details about numerically implementing the three approaches, emphasizing the choices practitioners face during implementation. Section 4.7 applies the three approaches to the model considered by Arias, Caldara and Rubio-Ramírez (2019). Sections 4.5–4.7 are self-contained. Section 4.8 concludes.

4.2 Robust Bayesian Analysis

4.2.1 Bayesian statistical decisions and inference

We start from the classical framework of statistical decision theory as in Wald (1950). Let $\mathbf{Y} \in \mathscr{Y} \subset \mathbb{R}^T$ be a sample whose probability distribution is assumed to belong to a parametric family of distributions $P_{\mathbf{Y}|\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{d_\theta}$. We denote the Borel σ-algebra of $\boldsymbol{\theta}$ by $\mathscr{B}(\Theta)$. Let $\delta(\cdot) : \mathscr{Y} \to \mathscr{A}$, $\delta \in \mathscr{D}$, be a nonrandomized statistical decision rule mapping a sample $Y \in \mathscr{Y}$ to an action $a \in \mathscr{A}$, where $\mathscr{D}$ is a set of (possibly constrained) decision rules and $\mathscr{A}$ is the set of actions the decision maker (DM) can take. Let $L(\boldsymbol{\theta}, a) : \Theta \times \mathscr{A} \to \mathbb{R}$ be the *loss* the DM incurs when the true parameter value is $\boldsymbol{\theta}$ and the action taken is a. The *risk* of the decision rule δ, denoted by $R(\boldsymbol{\theta}, \delta)$, measures the average loss under repeated sampling of the sample $\mathbf{Y} \sim P_{\mathbf{Y}|\boldsymbol{\theta}}$,

$$R(\boldsymbol{\theta}, \delta) \equiv E_{\mathbf{Y}|\boldsymbol{\theta}}(L(\boldsymbol{\theta}, \delta(\mathbf{Y}))) = \int_{\mathscr{Y}} L(\boldsymbol{\theta}, \delta(\mathbf{y})) dP_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{y}). \tag{4.1}$$

Since a decision rule dominating the others uniformly over $\boldsymbol{\theta}$ is usually not available, the ranking of decision rules depends on how the DM handles uncertainty about the unknown parameter $\boldsymbol{\theta}$.

If the DM could express the uncertainty about $\boldsymbol{\theta}$ in the form of a probability distribution $\pi_{\boldsymbol{\theta}}$ of the measurable space $(\Theta, \mathscr{B}(\Theta))$, they would choose δ that performs best in terms of *Bayes risk*

$$r(\pi_{\boldsymbol{\theta}}, \delta) \equiv \int_{\Theta} R(\boldsymbol{\theta}, \delta) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}). \tag{4.2}$$

This corresponds to the Bayesian decision principle, which is easy to derive and implement, and is favored by statisticians and decision theorists on the basis of the likelihood principle, conditionality viewpoint, and guaranteed admissibility (see, for example, the discussions in Chapters 1 and 4 of Berger (1985)). Under weak regularity conditions (e.g., Brown and Purves (1973)), the Bayes decision δ_{bayes} minimizing (4.2) can be obtained by minimizing the *posterior expected loss* $\rho(\pi_{\boldsymbol{\theta}}, a)$ at each realization of the sample supported by the marginal likelihood, $m(\cdot) = \int_{\Theta} P_{\mathbf{Y}|\boldsymbol{\theta}}(\cdot) d\pi_{\boldsymbol{\theta}}$. That is, $\delta_{Bayes}(\mathbf{y})$ minimizes in a

$$\rho(\pi_{\boldsymbol{\theta}}, a) \equiv \int_{\Theta} L(\boldsymbol{\theta}, a) d\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}), \tag{4.3}$$

where $\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\cdot|\mathbf{y})$ is the posterior (distribution) of $\boldsymbol{\theta}$ given the realization of the sample $\mathbf{Y} = \mathbf{y}$. If the goal of the analysis is to summarize uncertainty about the unknown parameter $\boldsymbol{\theta}$ upon observing the data, it suffices to report the posterior $\pi_{\boldsymbol{\theta}|\mathbf{Y}}$ or its summary statistics, which is feasible in many contexts thanks to advances in Monte Carlo sampling methods.

The Bayesian approach to statistical decision-making and inference is conceptually straightforward and easy to implement numerically as long as the DM specifies a triplet of loss function, likelihood and prior for $\boldsymbol{\theta}$. Specifying these inputs, however, can be a challenge in practice. The DM may not be sure about how to choose the loss function and/or the class of parametric distributions for the likelihood. Arguably, the dominant concern in Bayesian practice is how to organize the DM's belief for $\boldsymbol{\theta}$ (or the lack thereof) in terms of a prior. To cope with these concerns, robust Bayesian analysis allows for multiplicity in each of these inputs and assesses the set of posterior expected losses or Bayes risks spanned by the set of inputs. If the DM is interested in an optimal decision subject to the set of posterior losses or Bayes risks, the robust Bayesian literature has considered minimizing the upper bound of the posterior expected losses or Bayes risks. See Dey and Micheas (2000), Shyamalkumar (2000), and references therein for robust Bayesian analysis with multiple losses and likelihoods. The focus of this paper is on robust Bayesian analysis with multiple priors.

4.2.2 Robust Bayesian analysis with multiple priors

Let $\Pi_{\boldsymbol{\theta}}$ be a set of priors for $\boldsymbol{\theta}$. In the subjective robust Bayesian sense, $\Pi_{\boldsymbol{\theta}}$ represents ambiguity such that the DM considers any prior in $\Pi_{\boldsymbol{\theta}}$ plausible and cannot judge which one is more credible than the others.

To summarize the posterior uncertainty for $\boldsymbol{\theta}$, we update the set of priors $\Pi_{\boldsymbol{\theta}}$ based on the likelihood $P_{\mathbf{Y}|\boldsymbol{\theta}}$. One approach is prior-by-prior updating, which is often referred to as the full Bayesian updating rule.[2] It applies Bayes' rule to each prior in Π_θ to obtain the set of posteriors $\Pi_{\boldsymbol{\theta}|Y}$,

$$\Pi_{\boldsymbol{\theta}|\mathbf{Y}} \equiv \left\{ \pi_{\boldsymbol{\theta}|\mathbf{Y}}(\cdot) = \frac{\int_{\{\boldsymbol{\theta}\in\cdot\}} P_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{y}) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int_{\Theta} P_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{y}) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})} : \pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}} \right\}. \tag{4.4}$$

Given the set of posteriors, the analysis proceeds by reporting various posterior quantities. For instance, the *lower and upper posterior probabilities* for the hypothesis $\{\boldsymbol{\theta} \in A\}$ are the lower and upper bounds of $\pi_{\boldsymbol{\theta}|\mathbf{Y}}(A)$ on $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$; for an arbitrary measurable subset $A \in \mathscr{B}(\Theta)$, we have

$$\text{lower posterior probability for } \boldsymbol{\theta}\text{:} \quad \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A) \equiv \inf_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(A), \tag{4.5}$$

$$\text{upper posterior probability for } \boldsymbol{\theta}\text{:} \quad \pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(A) \equiv \sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(A) = 1 - \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A^c).$$

[2]The literature has considered different ways to update the set of priors. For example, the maximum likelihood updating rule considered and axiomatized by Gilboa and Schmeidler (1993) uses the observed sample to select a prior by maximizing the marginal likelihood and then applies Bayes' rule. This way of updating is known as Type-II maximum likelihood (Good (1965)) or, equivalently, as the empirical Bayes method (e.g., Robbins (1956), Berger and Berliner (1986)). For an axiomatization of the full Bayesian updating rule, see Pires (2002).

For example, $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A)$ can be interpreted as saying "the posterior credibility for the hypothesis $\{\boldsymbol{\theta} \in A\}$ is at least equal to $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A)$, no matter which prior in $\Pi_{\boldsymbol{\theta}}$ one assumes."

The corresponding probabilities for a parameter transformation $\eta = h(\boldsymbol{\theta}) \in \mathscr{H}$ are obtained as

$$\begin{aligned}
\text{lower posterior probability for } \eta = h(\boldsymbol{\theta})\text{:} \quad \pi_{\eta|\mathbf{Y}*}(D) &\equiv \inf_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta}) \in D), \\
\text{upper posterior probability for } \eta = h(\boldsymbol{\theta})\text{:} \quad \pi^*_{\eta|\mathbf{Y}}(D) &\equiv \sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \pi_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta}) \in D) \\
&= 1 - \pi_{\eta|\mathbf{Y}*}(D^c),
\end{aligned} \tag{4.6}$$

for any Borel set $D \subset \mathscr{H}$. If $\eta = h(\boldsymbol{\theta})$ is a scalar parameter of interest, quantities often reported in Bayesian global sensitivity analysis are the bounds for the posterior mean of η,

$$\left[\inf_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \int_{\Theta} h(\boldsymbol{\theta}) d\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}), \sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} \int_{\Theta} h(\boldsymbol{\theta}) d\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}) \right]. \tag{4.7}$$

In addition, the robust Bayesian counterpart of the highest posterior density region for η can be defined by a set $C_\alpha \subset \mathscr{H}$ such that the posterior lower probability is greater than or equal to α,

$$\pi_{\eta|\mathbf{Y}*}(C_\alpha) \geq \alpha. \tag{4.8}$$

Such C_α is interpreted as "a set on which the posterior credibility of η is at least α, *no matter which posterior is chosen within the set*". GK call C_α a *robust credible region with credibility* α.

The prior set $\Pi_{\boldsymbol{\theta}}$ can also generate the sets of Bayes risks and posterior expected losses; for any $\tilde{\pi}_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}$ and decision function $\delta \in \mathscr{D}$, it holds

$$\inf_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} r(\pi_{\boldsymbol{\theta}}, \delta) \leq r(\tilde{\pi}_{\boldsymbol{\theta}}, \delta) \leq \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} r(\pi_{\boldsymbol{\theta}}, \delta), \tag{4.9}$$

and for any action $a \in \mathscr{A}$ and sample realization $\mathbf{y} \in \mathscr{Y}$, it holds

$$\inf_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a) \leq \rho(\tilde{\pi}_{\boldsymbol{\theta}}, a) \leq \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a). \tag{4.10}$$

The set of posterior expected losses coincides with the set (4.7) with $h(\cdot)$ set to $L(\boldsymbol{\theta}, a)$.

Interpreting the set of priors as the DM's ambiguous belief, an unconditional optimality criterion attractive to the ambiguity-averse DM is the unconditional *Gamma minimax* criterion. It defines an optimal decision δ^* by minimizing the worst-case Bayes risk,

$$\delta^* \equiv \arg \inf_{\delta \in \mathscr{D}} \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} r(\pi_{\boldsymbol{\theta}}, \delta). \tag{4.11}$$

A similar but distinct optimality criterion is the *conditional Gamma minimax* criterion, which defines an optimal action by minimizing the worst-case posterior expected loss conditional on $\mathbf{y}$,

$$a^*_{\mathbf{y}} \equiv \arg \inf_{a \in \mathscr{A}} \sup_{\pi_{\boldsymbol{\theta}} \in \Pi_{\boldsymbol{\theta}}} \rho(\pi_{\boldsymbol{\theta}}, a). \tag{4.12}$$

The unconditional Gamma minimax decision δ^* and the conditional Gamma minimax action $a^*_{\mathbf{y}}$ do not generally agree. An advantage of the former is its guaranteed admissibility, since, under the regularity conditions leading to the minimax theorem, the Gamma minimax decision is Bayes optimal under a least-favorable prior. Hence, admissibility of δ^* follows if the least-favorable prior supports the whole of Θ or if δ^* is a unique Bayes decision under the least-favorable prior. On the other hand, obtaining δ^* is challenging analytically and numerically, unless we limit the analysis to simple models with a particular set of priors. See Chamberlain (2000) for an application to portfolio choice and Vidakovic (2000) for a review. The conditional Gamma minimax action is easier to analyze and implement as the minimax problem only involves action a. However, a potential downside is that this criterion is not guaranteed to give an admissible decision. DasGupta and Studden (1989) consider conditional Gamma minimax estimators for the normal means model. See also Betrò and Ruggeri (1992) for other examples.

4.2.3 Examples of sets of priors

We review different constructions of the prior set $\Pi_{\boldsymbol{\theta}}$ considered in the literature.

Example 4.2.1. In one the earliest applications of Bayesian global sensitivity analysis, Chamberlain and Leamer (1976) and Leamer (1978, 1982) consider the regression model

$$\underset{T\times 1}{\mathbf{y}} = \underset{T\times k}{\mathbf{X}}\underset{k\times 1}{\boldsymbol{\beta}} + \underset{T\times 1}{\boldsymbol{\varepsilon}}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_T(0, \sigma^2 \mathbf{I}_T),$$

and specify a set of conjugate priors where the variance of the conjugate Gaussian prior for $\boldsymbol{\beta}$ varies over a certain set. These works derive a closed-form representation for the set of posterior means of $\boldsymbol{\beta}$ and study its analytical properties.

Example 4.2.2 (ε-contaminated set). A well-studied set of priors is the ε-contaminated set of priors (e.g., Huber (1973), Berger (1984, 1985), Berger and Berliner (1986), Sivaganesan and Berger (1989)). Its canonical representation is

$$\Pi_{\boldsymbol{\theta}} = \{\pi_{\boldsymbol{\theta}} = (1-\varepsilon)\pi^0_{\boldsymbol{\theta}} + \varepsilon q_{\boldsymbol{\theta}} : q_{\boldsymbol{\theta}} \in \mathcal{Q}_{\boldsymbol{\theta}}\}, \tag{4.13}$$

where the inputs to be specified by the user are $\pi^0_{\boldsymbol{\theta}}$, the base prior representing a benchmark prior with limited confidence, $\varepsilon \in [0,1]$, the amount of contamination that gauges the uncertainty on the base prior, and $\mathcal{Q}_{\boldsymbol{\theta}}$, the set of contamination distributions which span the plausible priors. Huber (1973) considers a set of arbitrary distributions for $\mathcal{Q}_{\boldsymbol{\theta}}$ and derives a closed-form expression for the lower and upper posterior probabilities. Berger and Berliner (1986) consider the empirical Bayes posterior (Type-II maximum likelihood) with various sets of contamination distributions including unimodal and symmetric ones, and Sivaganesan and Berger (1989) derive the set of posteriors with contaminations preserving unimodality of $\pi_{\boldsymbol{\theta}}$.

Example 4.2.3 (Priors with fixed marginal). In the presence of multiple parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{d_\theta})$, $d_\theta \geq 2$, it is often feasible to elicit the marginal distribution of the prior for each component, while eliciting their dependence is difficult. Lavine et al. (1991) consider an ε-contaminated set of priors with $\mathscr{Q}_{\boldsymbol{\theta}}$ consisting of the set of priors sharing fixed marginals for each component in $\boldsymbol{\theta}$, while their dependence is unconstrained. They propose linearization techniques to solve the optimization in (4.7). Moreno and Cano (1995) fix the prior marginals of only a subset of the parameters in $\boldsymbol{\theta}$.

Example 4.2.4 (Priors known up to coarsened domain). Kudō (1967) and Manski (1981) study the set of posterior distributions and Gamma minimax statistical decisions when the analyst can elicit a prior distribution only up to a class of coarsened subsets of $\boldsymbol{\theta}$, i.e., a σ-algebra smaller than the Borel σ-algebra of $\boldsymbol{\theta}$, $\mathscr{B}(\Phi)$. Formally, consider a transformation $\boldsymbol{\phi} = g(\boldsymbol{\theta})$, where g is a many-to-one function $g : \Theta \to \Phi$ that coarsens the parameter space of $\boldsymbol{\theta}$, and $\boldsymbol{\phi}$ indexes the sets in the partition of $\boldsymbol{\theta}$. Given a unique prior $\pi_{\boldsymbol{\phi}}$ on $(\Phi, \mathscr{B}(\Phi))$, the set of priors for $\boldsymbol{\theta}$ consists of the distributions of $\boldsymbol{\theta}$ that imply the distribution of $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ is the given $\pi_{\boldsymbol{\phi}}$,

$$\Pi_{\boldsymbol{\theta}} = \{\pi_{\boldsymbol{\theta}} : \pi_{\boldsymbol{\theta}}(g^{-1}(B)) = \pi_{\boldsymbol{\phi}}(B),\ \forall B \in \mathscr{B}(\Phi)\}. \tag{4.14}$$

This set of priors can be interpreted as a special case of the general construction in Wasserman (1990). Wasserman (1990) considers a set of priors whose lower and upper probabilities are given by the containment and capacity functional of a random set $\Gamma : \Phi \rightrightarrows \Theta$, i.e., given $\pi_{\boldsymbol{\phi}}$, a probability measure on $(\Phi, \mathscr{B}(\Phi))$,

$$\Pi_{\boldsymbol{\theta}} = \left\{\pi_{\boldsymbol{\theta}} : \pi_{\boldsymbol{\theta}*}(A) \leq \pi_{\boldsymbol{\theta}}(A) \leq \pi^*_{\boldsymbol{\theta}}(A),\ \forall A \in \mathscr{B}(\Theta)\right\}, \tag{4.15}$$

where $\pi_{\boldsymbol{\theta}*}(A) = \pi_{\boldsymbol{\phi}}(\Gamma(\boldsymbol{\phi}) \subset A)$ and $\pi^*_{\boldsymbol{\theta}}(A) = \pi_{\boldsymbol{\phi}}(\Gamma(\boldsymbol{\phi}) \cap A \neq \emptyset)$. In view of random set theory, this set of priors can be interpreted as the set of selectionable distributions from the random set $\Gamma(\boldsymbol{\phi})$, $\boldsymbol{\phi} \sim \pi_{\boldsymbol{\phi}}$. See **?**, Molchanov (2005), and Molchanov and Molinari (2018). This set of priors coincides with (4.14) in the special case where $\Gamma(\boldsymbol{\phi})$ is set to $g^{-1}(\boldsymbol{\phi})$. Wasserman (1990) derives analytically the lower and upper posterior probabilities when $\Pi_{\boldsymbol{\theta}}$ is given in the form (4.15).

Example 4.2.5 (Priors in an information neighborhood). In the minimax approaches to robust estimation and robust control, the set of distributions one wishes to be robust against is formed by an information neighborhood around a benchmark distribution $\pi_{\boldsymbol{\theta}}^0$. One can consider a variety of statistical divergence criteria to define the neighborhood, including the KL divergence and the Hellinger distance; see, for example, Peterson et al. (2000), Hansen and Sargent (2001) and Kitamura et al. (2013). This approach offers a flexible and analytically tractable way to define the set of priors for robust Bayesian analysis. Along this line, Ho (2020) introduces the set of priors through the KL neighborhood centered at a benchmark prior $\pi_{\boldsymbol{\theta}}^0$,

$$\Pi_{\boldsymbol{\theta}} = \left\{ \pi_{\boldsymbol{\theta}} : \int_{\Theta} \log \left(\frac{d\pi_{\boldsymbol{\theta}}}{d\pi_{\boldsymbol{\theta}}^0} \right) d\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \leq \lambda \right\}, \tag{4.16}$$

where $\lambda > 0$ is the radius of the KL neighborhood specified by the user. Watson and Holmes (2016) analyze posterior Gamma minimax actions with a KL neighborhood around the benchmark posterior. As discussed in Ho (2020), a convenient feature of this approach is that it is easy to additionally impose moment constraints on the priors and/or posteriors.

4.3 Robust Bayesian Analysis for Set-identified Models

This section discusses how the general framework of robust Bayesian analysis introduced in the previous section can be extended to a class of set-identified structural models.

4.3.1 Set-identified structural models

Non-identification of a structural parameter $\boldsymbol{\theta}$ arises when multiple values of $\boldsymbol{\theta}$ are observationally equivalent; that is, there exist $\boldsymbol{\theta}$ and $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ such that $p(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta}')$ for every $\mathbf{y} \in \mathscr{Y}$ (Rothenberg (1971)). Observational equivalence can be represented by a many-to-one function $g : (\Theta, \mathscr{A}) \to (\Phi, \mathscr{B})$, such that $g(\boldsymbol{\theta}) = g(\boldsymbol{\theta}')$ if and only if $p(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta}')$ for all $\mathbf{y} \in \mathscr{Y}$ (e.g., Barankin (1960)). This relationship partitions the parameter space Θ into equivalent classes, in each of which the likelihood of $\boldsymbol{\theta}$ is "flat" irrespective of observations, and $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ maps each of the equivalent classes to a point in a parameter space Φ. Following the terminology of structural models in econometrics (Koopmans and Reiersol (1950)), $\boldsymbol{\phi} = g(\boldsymbol{\theta})$ is the reduced-form parameter indexing the distribution of the data. The likelihood depends on $\boldsymbol{\theta}$ only through $\boldsymbol{\phi} = g(\boldsymbol{\theta})$; that is, there exists a $\mathscr{B}(\Phi)$-measurable function $\hat{p}(\mathbf{y}|\cdot)$ such that $p(\mathbf{y}|\boldsymbol{\theta}) = \hat{p}(\mathbf{y}|g(\boldsymbol{\theta}))$ for every $\mathbf{y} \in \mathscr{Y}$ and $\boldsymbol{\theta} \in \Theta$.

The identified set of $\boldsymbol{\theta}$ is the inverse image of $g(\cdot)$: $IS_{\theta}(\boldsymbol{\phi}) = \{\boldsymbol{\theta} \in \Theta : g(\boldsymbol{\theta}) = \boldsymbol{\phi}\}$, where $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ and $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}')$ for $\boldsymbol{\phi} \neq \boldsymbol{\phi}'$ are disjoint and $\{IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}) : \boldsymbol{\phi} \in \Phi\}$ is a partition of Θ. For the parameter of interest $\eta = h(\boldsymbol{\theta})$ with $h : \Theta \to \mathscr{H}$, $\mathscr{H} \subset \mathbb{R}^{d_\eta}$, $d_\eta < \infty$, we define the identified set as the projection of $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ onto $\mathscr{H}$ through $h(\cdot)$, $IS_\eta(\boldsymbol{\phi}) \equiv \{h(\boldsymbol{\theta}) : \boldsymbol{\theta} \in IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})\}$. The parameter $\eta = h(\boldsymbol{\theta})$ is point- or set-identified at $\boldsymbol{\phi}$ if $IS_\eta(\boldsymbol{\phi})$ is a singleton or not a singleton, respectively. By the definition of observational equivalence, $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ and $IS_\eta(\boldsymbol{\phi})$ are the sharp identification regions at every distribution of data indexed by $\boldsymbol{\phi}$.

In SVARs with sign restrictions, the model can be observationally restrictive in the sense of Koopmans and Reiersol (1950). This means the model is falsifiable and $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ can be empty for some $\boldsymbol{\phi} \in \Phi$ on which the reduced-form likelihood is well defined.

4.3.2 Influence of prior choice under set-identification

Let $\pi_{\boldsymbol{\theta}}$ be a prior for $\boldsymbol{\theta}$ and $\pi_{\boldsymbol{\phi}}$ be the corresponding prior for $\boldsymbol{\phi}$ induced by $\pi_{\boldsymbol{\theta}}$ and $g(\cdot)$:

$$\pi_{\boldsymbol{\phi}}(B) = \pi_{\boldsymbol{\theta}}(IS_{\boldsymbol{\theta}}(B)) \quad \text{for all } B \in \mathscr{B}(\Phi). \tag{4.17}$$

From the definition of the reduced-form parameter, the likelihood for $\boldsymbol{\theta}$ is flat on $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$ for any $\mathbf{y}$, which implies conditional independence $\boldsymbol{\theta} \perp \mathbf{Y}|\boldsymbol{\phi}$. Hence, as obtained by Poirier (1998), Moon and Schorfheide (2012), and Baumeister and Hamilton (2015), the posterior of $\boldsymbol{\theta}$, $\pi_{\boldsymbol{\theta}|\mathbf{Y}}$, can be expressed as

$$\pi_{\boldsymbol{\theta}|\mathbf{Y}}(A) = \int_{\Phi} \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(A) d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \quad A \in \mathscr{B}(\Theta), \tag{4.18}$$

where $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ is the conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\phi}$ whose support agrees with or is contained in $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$, and $\pi_{\boldsymbol{\phi}|\mathbf{Y}}$ is the posterior of $\boldsymbol{\phi}$. This expression shows that the prior of the reduced-form parameter, $\pi_{\boldsymbol{\phi}}$, can be updated by the data, whereas the conditional prior of $\boldsymbol{\theta}$ given $\boldsymbol{\phi}$ is never updated because the likelihood is flat on $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi}) \subset \Theta$ for any realization of the sample. In this sense, one can interpret $\pi_{\boldsymbol{\phi}}$ as the *revisable prior knowledge* and the conditional priors, $\left\{\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\cdot|\boldsymbol{\phi}) : \boldsymbol{\phi} \in \Phi\right\}$, as the *unrevisable prior knowledge*.

Marginalizing $\pi_{\boldsymbol{\theta}|\mathbf{Y}}$ for $\eta = h(\boldsymbol{\theta})$ gives

$$\pi_{\eta|\mathbf{Y}}(D) = \int_{\Phi} \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(h(\boldsymbol{\theta}) \in D) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} = \int_{\Phi} \pi_{\eta|\boldsymbol{\phi}}(D) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} \tag{4.19}$$

for $D \in \mathscr{B}(\mathscr{H})$, where $\pi_{\eta|\boldsymbol{\phi}}$ is the conditional prior for η given $\boldsymbol{\phi}$, which is by construction supported on $IS_{\eta}(\boldsymbol{\phi})$. If η is set-identified, its posterior thus has a non-degenerate unrevisable component $\pi_{\eta|\boldsymbol{\phi}}$.

The previous discussion clarifies the following features of Bayesian inference under set-identification:

1. As discussed in Poirier (1998), the lack of identification for η does not mean the prior for η is not updated by the data. The prior for η can be updated, but this happens only through the update of the prior for the reduced-form parameter $\boldsymbol{\phi}$. Comparing the prior and posterior for η therefore does not indicate whether or not the parameter is point-identified.

2. Since the posteriors for $\boldsymbol{\theta}$ and η involve nonrevisable priors, they are sensitive to the choice of prior even asymptotically. In particular, the posterior for η is sensitive to perturbations of the prior that change the shape of $\pi_{\eta|\boldsymbol{\phi}}$. This suggests that posterior sensitivity, rather than the comparison of the shapes of prior and posterior, is informative about the strength of identification. This feature is similar to the local sensitivity analysis in Müller (2012).

3. The reduced-form parameter $\boldsymbol{\phi}$ is identified by construction. If the likelihood for $\boldsymbol{\phi}$ converges to its true value $\boldsymbol{\phi}_0$ in large samples, the posterior for η converges to the conditional prior $\pi_{\eta|\boldsymbol{\phi}}$ given $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. Since the support of

$\pi_{\eta|\boldsymbol{\phi}}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ is equal to or contained in $IS_\eta(\boldsymbol{\phi}_0)$, the asymptotic posterior does not lead to an estimate for η lying outside of its identified set. However, the shape of the asymptotic posterior on $IS_\eta(\boldsymbol{\phi}_0)$ is determined entirely by the prior.

4.3.3 Full ambiguity for the unrevisable prior

The discussion in the previous subsection motivates a key feature of the robust Bayesian approaches for set-identified models that we discuss in this paper: they assume multiple priors for the unrevisable component of the prior (the prior for the structural parameter $\boldsymbol{\theta}$ given the reduced-form parameter $\boldsymbol{\phi}$), but maintain a single prior for the revisable component (the prior for $\boldsymbol{\phi}$).[3]

In this section we review the full-ambiguity approach of GK for general models, but one can consider a variety of approaches to refine the set of priors in GK to reflect partial prior knowledge about the unrevisable component of the prior. Examples include the ε-contaminated set of priors (Example 4.2.2) and the KL-neighborhood set of priors (Example 4.2.5). For SVARs, GKV investigate the former and GKU investigate a variation of the latter, as we review in Section 4.5 below.

GK construct a set of priors for $\boldsymbol{\theta}$ constrained by a single proper prior $\pi_{\boldsymbol{\phi}}$ for $\boldsymbol{\phi} = g(\boldsymbol{\theta})$, supported on $g(\Theta)$,

$$\Pi_{\boldsymbol{\theta}}^{FA} \equiv \{\pi_{\boldsymbol{\theta}} : \pi_{\boldsymbol{\theta}}(IS_{\boldsymbol{\theta}}(B)) = \pi_{\boldsymbol{\phi}}(B),\ \forall B \in \mathscr{B}(\Phi)\}, \tag{4.20}$$

where $IS_{\boldsymbol{\theta}}(B) = \cup_{\boldsymbol{\phi}\in B} IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$. Noting $IS_{\boldsymbol{\theta}}(\cdot) = g^{-1}(\cdot)$, this takes the form as the set of priors in (4.14). An equivalent but perhaps more intuitive way to introduce ambiguity for the unrevisable prior is in terms of the set of conditional priors:

$$\Pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}^{FA} \equiv \left\{\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})) = 1, \pi_{\boldsymbol{\phi}} - \textit{almost surely}\right\}. \tag{4.21}$$

$\Pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}^{FA}$ consists of arbitrary conditional priors as long as they assign probability

[3]GK discuss how this is motivated by the asymptotically negligible effect of the prior choice for $\boldsymbol{\phi}$ and by a desire to avoid possible issues of non-convergence of the set of posteriors. In addition, for $\boldsymbol{\phi}$ one can apply existing methods for constructing a non-informative prior such as Jeffreys' prior or for selecting a data-driven prior.

one to the identified set of $\boldsymbol{\theta}$, and is linked to $\Pi^{FA}_{\boldsymbol{\theta}}$ in (4.20) by $\Pi^{FA}_{\boldsymbol{\theta}} = \{\pi_{\boldsymbol{\theta}}(\cdot) = \int \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\cdot) d\pi_{\boldsymbol{\phi}}(\boldsymbol{\phi}) : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}\}$.

Applying Bayes' rule to each prior in $\Pi^{FA}_{\boldsymbol{\theta}}$ gives a set of posteriors for $\boldsymbol{\theta}$. Marginalizing each posterior and invoking (4.19) and (4.21) generates a set of posteriors for the parameter of interest η,

$$\Pi^{FA}_{\eta|\mathbf{Y}} \equiv \left\{ \pi_{\eta|\mathbf{Y}}(\cdot) = \int_{\Phi} \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(h(\boldsymbol{\theta}) \in \cdot) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \right\}. \tag{4.22}$$

Under mild regularity conditions (Assumption 1 in GK), GK derive the lower and upper posterior probabilities of $\Pi^{FA}_{\eta|\mathbf{Y}}$ as

$$\pi_{\eta|\mathbf{Y}*}(D) = \pi_{\boldsymbol{\phi}|\mathbf{Y}}(\{\boldsymbol{\phi} : IS_{\eta}(\boldsymbol{\phi}) \subset D\}), \quad \pi^*_{\eta|\mathbf{Y}}(D) = \pi_{\boldsymbol{\phi}|\mathbf{Y}}(\{\boldsymbol{\phi} : IS_{\eta}(\boldsymbol{\phi}) \cap D \neq \emptyset\}), \tag{4.23}$$

for $D \in \mathscr{B}(\mathscr{H})$, and show that the set of posterior probabilities $\{\pi_{\eta|\mathbf{Y}}(D) : \pi_{\eta|\mathbf{Y}} \in \Pi^{FA}_{\eta|\mathbf{Y}}\}$ coincides with the connected intervals $[\pi_{\eta|\mathbf{Y}*}(D), \pi^*_{\eta|\mathbf{Y}}(D)]$, which implies that any posterior probability in this set can be attained by some posterior in $\Pi^{FA}_{\eta|\mathbf{Y}}$. The expression for $\pi_{\eta|\mathbf{Y}*}(D)$ shows that the lower probability on D is the probability that the (random) identified set $IS_{\eta}(\boldsymbol{\phi})$ is contained in D in terms of the posterior probability of $\boldsymbol{\phi}$. The upper probability is the probability that the identified set hits D. These closed-form expressions of the lower and upper probabilities suggest how to compute them in practice. For instance, to approximate $\pi_{\eta|\mathbf{Y}*}(D)$, one obtains Monte Carlo draws of $\boldsymbol{\phi}$ from its posterior and computes the proportion of the draws satisfying $IS_{\eta}(\boldsymbol{\phi}) \subset D$.

GK show that the set of posterior means of η coincides with the Aumann expectation of the convex hull of the identified set, $co(IS_{\eta}(\boldsymbol{\phi}))$, with respect to $\pi_{\boldsymbol{\phi}|Y}$. In particular, if η is a scalar and denoting the convexified identified set for η by $[\ell(\boldsymbol{\phi}), u(\boldsymbol{\phi})] = co(IS_{\eta}(\boldsymbol{\phi}))$, the set of posterior means for η is the interval connecting the posterior means of $\ell(\boldsymbol{\phi})$ and $u(\boldsymbol{\phi})$,

$$\left\{ E_{\eta|\mathbf{Y}}(\eta) : \pi_{\eta|\mathbf{Y}} \in \Pi^{FA}_{\eta|\mathbf{Y}} \right\} = \left[E_{\boldsymbol{\phi}|\mathbf{Y}}(\ell(\boldsymbol{\phi})), E_{\boldsymbol{\phi}|\mathbf{Y}}(u(\boldsymbol{\phi})) \right]. \tag{4.24}$$

The set of posterior τ-th quantiles of η can be computed by first applying (4.23) with $D = (-\infty, t]$, $-\infty < t < \infty$ to obtain the set of the posterior cumulative distribution functions of η for each t and then inverting the upper and lower bounds of this set at $\tau \in (0,1)$.

Given the representation of the lower posterior probability (4.23), a robust credible region satisfying (4.8) with credibility $\alpha \in (0,1)$ can be expressed as

$$\pi_{\eta|\mathbf{Y}*}(C_\alpha) = \pi_{\boldsymbol{\phi}|\mathbf{Y}}(IS_\eta(\boldsymbol{\phi}) \subset C_\alpha)) \geq \alpha. \tag{4.25}$$

GK propose to report the smallest robust credible region (i.e., C_α with the smallest volume):

$$C_\alpha^* \in \arg\min_{C \in \mathscr{C}} Leb(C), \text{ s.t. } \pi_{\phi|\mathbf{Y}}(IS_\eta(\boldsymbol{\phi}) \subset C)) \geq \alpha, \tag{4.26}$$

where $Leb(C)$ is the volume of C in terms of the Lebesgue measure and $\mathscr{C}$ is a family of subsets in $\mathscr{H}$. The credible regions for the identified set proposed in Moon and Schorfheide (2011), Norets and Tang (2014) and Kline and Tamer (2016) satisfy (4.25), so they can be interpreted as robust credible regions, but they are not optimized in terms of volume.

4.4 Analytical Results for Set of Posterior Moments

In this section we present new and general theoretical results linking the set of posterior moments in (4.7) and the lower and upper posterior probabilities in (4.5). These results are useful because: 1) they provide a general approach to deriving and computing the set of posterior moments for sensitivity analysis; and 2) they help solve the posterior Gamma minimax problem.

The lower and upper posterior probabilities viewed as functions of $A \in \mathscr{B}(\Theta)$ are nonnegative and monotone set functions ($0 \leq \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A_1) \leq \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(A_2)$ for $A_1 \subset A_2$), while they are non-additive; a measure μ defined on $\mathscr{B}(\Theta)$ is non-additive if $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) \neq \mu(A_1) + \mu(A_2)$ for some $A_1 \neq A_2$, $A_1, A_2 \in \mathscr{B}(\Theta)$. A non-additive measure μ is called *submodular* or *2-alternating* if

$$\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) \leq \mu(A_1) + \mu(A_2), \ \forall A_1, A_2 \in \mathscr{B}(\Theta). \tag{4.27}$$

If the inequality in (4.27) is reversed, μ is called *supermodular* or *2-monotone*. The *core* of a non-additive measure μ is defined by

$$core(\mu) \equiv \{\pi \text{ probability measure } : \pi(A) \geq \mu(A) \text{ holds for all } A \in \mathscr{B}(\Theta)\}. \tag{4.28}$$

The next condition concerns the supermodular (submodular) property of the lower (upper) posterior probability and the richness of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ in the sense that $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ agrees with the core of its lower probability. The latter property is called representability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ by a lower probability (Huber (1973)).

Condition 4.4.1. *The set of posteriors $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ satisfies the following two conditions:*

(i) The lower probability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$, $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}$, is supermodular, or equivalently, the upper probability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$, $\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}$, is submodular.

(ii) $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ is representable by its lower probability $\pi_{\boldsymbol{\theta}|\mathbf{Y}*}$, i.e., $\Pi_{\boldsymbol{\theta}|\mathbf{Y}} = core(\pi_{\boldsymbol{\theta}|\mathbf{Y}*})$ holds.

Under Condition 4.4.1, we obtain the following result expressing the bounds of the posterior mean of $h(\boldsymbol{\theta})$ on $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$ in terms of the Choquet expectation of $h(\boldsymbol{\theta})$ with respect to the upper probability of $\Pi_{\boldsymbol{\theta}|\mathbf{Y}}$. This result follows directly from Proposition 10.3 of Denneberg (1994), so we omit a proof.

Theorem 4.4.1. *Let $h : \Theta \to \mathbb{R}$ be a measurable real-valued function. If Condition 4.4.1 holds, then*

$$\sup_{\pi_{\boldsymbol{\theta}|\mathbf{Y}} \in \Pi_{\boldsymbol{\theta}|\mathbf{Y}}} E_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta})) = \int h(\boldsymbol{\theta}) d\pi^*_{\boldsymbol{\theta}|\mathbf{Y}} \tag{4.29}$$

$$\inf_{\pi_{\boldsymbol{\theta}|Y} \in \Pi_{\boldsymbol{\theta}|Y}} E_{\boldsymbol{\theta}|\mathbf{Y}}(h(\boldsymbol{\theta})) = -\int (-h(\boldsymbol{\theta})) d\pi^*_{\boldsymbol{\theta}|\mathbf{Y}},$$

where the integral with respect to non-additive measure $\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}$ is defined as the Cho-

quet integral, i.e., for measurable real-valued function $f : \Theta \to \mathbb{R}$,

$$\int f(\boldsymbol{\theta}) d\pi^*_{\boldsymbol{\theta}|\mathbf{Y}} \equiv \int_0^\infty \pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\})dt + \int_{-\infty}^0 [\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\}) - 1]dt \tag{4.30}$$

$$= \int_0^\infty \pi^*_{\boldsymbol{\theta}|\mathbf{Y}}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\})dt - \int_{-\infty}^0 \pi_{\boldsymbol{\theta}|\mathbf{Y}*}(\{\boldsymbol{\theta} : f(\boldsymbol{\theta}) \geq t\})dt. \tag{4.31}$$

In many cases, it is not straightforward to check whether Condition 4.4.1 holds. One important case in which this condition is guaranteed to hold is for a set of posteriors whose lower probability can be represented as the containment probability of some random set and the set of posteriors includes any measurable selections of the random set. The set of posteriors for set-identified models considered by GK satisfies this condition. This claim follows from the fact that the posterior lower probability of $\Pi^{FA}_{\boldsymbol{\theta}|\mathbf{Y}}$ in (4.23) is the containment probability of the random set $IS_{\boldsymbol{\theta}}(\boldsymbol{\phi})$, $\boldsymbol{\phi} \sim \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}$, and from Artstein's inequality (**?**) for selectionable distributions from the random set. Hence, Theorem 4.4.1 always applies to robust Bayesian analysis with the GK set of priors.

The equivalence relationship (4.29) in Theorem 4.4.1 is valid for general posterior sets as far as they satisfy Condition 4.4.1, and it is not limited to the GK set of priors for set-identified models. Setting $\eta = h(\boldsymbol{\theta})$ in this theorem gives the upper and lower bounds of the posterior mean of η in terms of the Choquet integral. If the Choquet integral (or the upper posterior probability $\pi^*_{\boldsymbol{\theta}|\mathbf{Y}}$) is analytically or numerically tractable in a given context, this theorem offers a general approach to deriving and computing the set of posterior moments.

Theorem 4.4.1 is also useful for solving the posterior Gamma minimax problem. The theorem implies that the worst-case posterior expected loss coincides with the Choquet expectation with respect to $\pi^*_{\eta|\mathbf{Y}}$. Furthermore, for the GK set of priors, the Choquet expectation of the loss given action $a \in \mathscr{A}$ can be expressed as (e.g., Theorem 5.1 in Molchanov (2005))

$$\begin{aligned}\int L(\eta,a)d\pi^*_{\eta|\mathbf{Y}} &= \int_0^\infty \pi_{\boldsymbol{\phi}|\mathbf{Y}}\left(\{\boldsymbol{\phi}:\{\eta:L(\eta,a)\geq t\}\cap IS_\eta(\boldsymbol{\phi})\neq\emptyset\}\right)dt \\ &= \int_0^\infty \pi_{\boldsymbol{\phi}|\mathbf{Y}}\left(\left\{\boldsymbol{\phi}:\sup_{\eta\in IS_\eta(\boldsymbol{\phi})}\{L(\eta,a)\}\geq t\right\}\right)dt \\ &= \int_\Phi \sup_{\eta\in IS_\eta(\boldsymbol{\phi})} L(\eta,a)d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}).\end{aligned}$$

We hence obtain the following theorem for the representation of the posterior Gamma minimax criterion:

Theorem 4.4.2. *Let $L(\eta,a)$ be a nonnegative loss function (e.g., the quadratic loss $L(\eta,a)=(\eta-a)^2$) and $\Pi_{\boldsymbol{\theta}}$ be the set of priors constructed in (4.20). With the set of posteriors $\Pi^{FA}_{\eta|\mathbf{Y}}$ obtained in (4.22), the upper posterior expected loss at action a satisfies*

$$\sup_{\pi_{\boldsymbol{\theta}}\in\Pi^{FA}_{\boldsymbol{\theta}}}\rho(\pi_{\boldsymbol{\theta}},a)=\int L(\eta,a)d\pi^*_{\eta|\mathbf{Y}}=\int_\Phi \sup_{\eta\in IS_\eta(\boldsymbol{\phi})} L(\eta,a)d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \tag{4.32}$$

provided the Choquet integral is finite, $\int L(\eta,a)d\pi^*_{\eta|\mathbf{Y}}(\eta)<\infty$.

The posterior Gamma minimax criterion shown in (4.32) combines the ambiguity about η (given $\boldsymbol{\phi}$, what we know about η is only that it lies within the identified set $IS_\eta(\boldsymbol{\phi})$) with the posterior uncertainty about the identified set $IS_\eta(\boldsymbol{\phi})$ (in finite samples, the identified set of η is known with some uncertainty as summarized by the posterior of $\boldsymbol{\phi}$). Since Theorem 4.4.2 imposes no assumption on the loss function other than its nonnegativity, this result is also applicable to a planner's policy decision problem under a set-identified social welfare criterion. See Manski (2000) for statistical decision theory applied to treatment choice under a set-identified welfare criterion.

Theorem 4.4.2 also suggests a simple numerical algorithm for computing the posterior Gamma minimax action using a Monte Carlo sample of $\boldsymbol{\phi}$ from its posterior $\pi_{\boldsymbol{\phi}|\mathbf{Y}}$. Let $\{\boldsymbol{\phi}_s\}_{s=1}^S$ be S random draws of $\boldsymbol{\phi}$ from the posterior $\pi_{\boldsymbol{\phi}|\mathbf{Y}}$. Then, the posterior Gamma minimax action $\hat{a}^*\in\arg\min_a\left\{\sup_{\pi_{\boldsymbol{\theta}}\in\Pi^{FA}_{\boldsymbol{\theta}}}\rho(\pi_{\boldsymbol{\theta}},a)\right\}$ can be approximated by

$$\hat{a}^* \in \arg\min_a \frac{1}{S}\sum_{s=1}^{S} \sup_{\eta \in IS_\eta(\boldsymbol{\phi}_s)} L(\eta, a).$$

The posterior Gamma minimax action does not generally coincide with an unconditional optimal Gamma minimax decision. This is also the case with the prior set (4.20), implying that $\hat{a}^*$ fails to be a Bayesian action with respect to any single prior in the set.

4.5 Robust Bayesian Inference in SVARs

This section discusses the approaches to conducting robust Bayesian analysis in SVARs in GK, GKV and GKU. We first describe the SVAR framework and outline some commonly used identifying restrictions. We assume the parameter of interest is an individual impulse response, but the approaches easily extend to other parameters, such as forecast error variance decompositions.

4.5.1 Setup

Consider an SVAR(p) for the n-dimensional vector $\mathbf{y}_t$:

$$\mathbf{A}_0\mathbf{y}_t = \sum_{j=1}^{p} \mathbf{A}_j\mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{n\times 1}, \mathbf{I}_n), \text{ for } t = 1,\ldots,T, \tag{4.33}$$

where $\mathbf{A}_0$ is invertible. The reduced-form VAR(p) model is

$$\mathbf{y}_t = \sum_{j=1}^{p} \mathbf{B}_j\mathbf{y}_{t-j} + \mathbf{u}_t, \tag{4.34}$$

where $\mathbf{B}_j = \mathbf{A}_0^{-1}\mathbf{A}_j$, $\mathbf{u}_t = \mathbf{A}_0^{-1}\boldsymbol{\varepsilon}_t$, and $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t') \equiv \boldsymbol{\Sigma} = \mathbf{A}_0^{-1}\left(\mathbf{A}_0^{-1}\right)'$. The reduced-form parameter is $\boldsymbol{\phi} = (\mathrm{vec}(\mathbf{B})', \mathrm{vech}(\boldsymbol{\Sigma})')' \in \boldsymbol{\Phi}$, where $\mathbf{B} = [\mathbf{B}_1,\ldots,\mathbf{B}_p]$. Let $\mathbf{Y}$ denote the sample.

In SVAR applications, we can set the structural parameter vector $\boldsymbol{\theta}$ considered in the general framework above as $\boldsymbol{\theta} = \left(\boldsymbol{\phi}', \mathrm{vec}(\mathbf{Q})'\right)'$, where $\mathbf{Q}$ is an $n \times n$ orthonormal matrix in the set $\mathscr{O}(n)$ of orthonormal matrices (e.g., Uhlig (2005) and Rubio-Ramírez et al. (2010)). $\boldsymbol{\theta}$ transforms the SVAR structural parameters $[\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_p]$ as $\mathbf{B} = \mathbf{A}_0^{-1}[\mathbf{A}_1,\ldots,\mathbf{A}_p]$, $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}\left(\mathbf{A}_0^{-1}\right)'$ and $\mathbf{Q} = \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{A}_0^{-1}$, where

$\mathbf{\Sigma}_{tr}$ is the lower-triangular Cholesky factor of $\mathbf{\Sigma}$ with nonnegative diagonal elements. This transformation is one-to-one and can be inverted as $\mathbf{A}_0 = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}$ and $[\mathbf{A}_1, \ldots, \mathbf{A}_p] = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}\mathbf{B}$.

We assume that the reduced-form VAR(p) model can be inverted into a VMA(∞) model:

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{u}_{t-j} = \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{\Sigma}_{tr} \mathbf{Q} \boldsymbol{\varepsilon}_{t-j},$$

where $\mathbf{C}_j$ is the j-th coefficient matrix of $\left(\mathbf{I}_n - \sum_{j=1}^{p} \mathbf{B}_j L^j\right)^{-1}$.

The h-th horizon impulse response is the $n \times n$ matrix $\mathbf{IR}^h$,

$$\mathbf{IR}^h = \mathbf{C}_h \mathbf{\Sigma}_{tr} \mathbf{Q} \quad , h = 0, 1, 2, \ldots, \tag{4.35}$$

and the long-run cumulative impulse-response matrix is

$$\mathbf{CIR}^\infty = \sum_{h=0}^{\infty} \mathbf{IR}^h = \left(\sum_{h=0}^{\infty} \mathbf{C}_h\right) \mathbf{\Sigma}_{tr} \mathbf{Q} = \left(\mathbf{I}_n - \sum_{j=1}^{p} \mathbf{B}_j\right)^{-1} \mathbf{\Sigma}_{tr} \mathbf{Q}. \tag{4.36}$$

The scalar parameter of interest η is the impulse-response:

$$\eta = IR_{ij}^h \equiv \mathbf{e}_i' \mathbf{C}_h \mathbf{\Sigma}_{tr} \mathbf{Q} \mathbf{e}_j \equiv \mathbf{c}_{ih}'(\boldsymbol{\phi})\, \mathbf{q}_j, \tag{4.37}$$

where $\mathbf{e}_i$ is the i-th column of $\mathbf{I}_n$, $\mathbf{c}_{ih}'(\boldsymbol{\phi})$ is the i-th row of $\mathbf{C}_h \mathbf{\Sigma}_{tr}$ and $\mathbf{q}_j$ is the j-th column of $\mathbf{Q}$.

4.5.2 Set-identifying restrictions in SVARs

An SVAR without identifying restrictions is set-identified because there are multiple values of $\mathbf{A}_0$ consistent with $\boldsymbol{\phi}$: $\left\{\mathbf{A}_0 = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1} : \mathbf{Q} \in \mathscr{O}(n)\right\}$. Imposing zero and/or sign restrictions can be viewed as constraining the set of orthonormal matrices to lie in a subspace $\mathscr{Q}(\boldsymbol{\phi})$ of $\mathscr{O}(n)$, which in turn yields the following identified set for the impulse-response η:

$$IS_\eta(\boldsymbol{\phi}) = \left\{\eta(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi})\right\}. \tag{4.38}$$

We now characterize the subspace $\mathcal{Q}(\boldsymbol{\phi})$ under typical zero restrictions in SVARs, under restrictions induced by external instruments in proxy SVARs and under sign restrictions. In addition to any such restrictions, we follow the convention in the literature and assume one always imposes the sign normalization restrictions $\operatorname{diag}(\mathbf{A}_0) = \operatorname{diag}\left(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\right) \geq \mathbf{0}_{n\times 1}$, which imply that a positive value of ε_{it} is a positive shock to the i-th equation of the SVAR at time t.

4.5.2.1 Zero restrictions

Commonly used zero restrictions in SVARs can be written as linear constraints on the columns of $\mathbf{Q}$. For example:

$$\begin{aligned}
((i,j)\text{-th element of } \mathbf{A}_0) &= 0 \iff \left(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_j\right)'\mathbf{q}_i = 0, \qquad (4.39)\\
\left((i,j)\text{-th element of } \mathbf{A}_0^{-1}\right) &= 0 \iff \left(\mathbf{e}_i'\boldsymbol{\Sigma}_{tr}\right)\mathbf{q}_j = 0,\\
((i,j)\text{-th element of } \mathbf{CIR}^{\infty}) &= 0 \iff \left[\mathbf{e}_i'\left(\mathbf{I}_n - \sum_{j=1}^{p}\mathbf{B}_j\right)^{-1}\boldsymbol{\Sigma}_{tr}\right]\mathbf{q}_j = 0.
\end{aligned}$$

We represent a collection of zero restrictions as:

$$F(\boldsymbol{\phi},\mathbf{Q}) \equiv \begin{pmatrix} F_1(\boldsymbol{\phi})\mathbf{q}_1 \\ F_2(\boldsymbol{\phi})\mathbf{q}_2 \\ \vdots \\ F_n(\boldsymbol{\phi})\mathbf{q}_n \end{pmatrix} = \mathbf{0}_{\sum_{i=1}^{n} f_i \times 1}, \qquad (4.40)$$

with $F_i(\boldsymbol{\phi})$ an $f_i \times n$ matrix. We assume that the imposed zero restrictions satisfy $f_i \leq n-i$, $i = 1,\ldots,n$, so we rule out over-identifying zero restrictions and SVARs that are locally identified but not globally identified (Bacchiocchi and Kitagawa (2020)). To facilitate computing identified sets, we adopt the following ordering convention for the variables in $\mathbf{y}_t$. Assume that the variables are ordered so the number of zero restrictions f_i imposed on the i-th column of $\mathbf{Q}$ satisfies $f_1 \geq f_2 \geq \cdots \geq f_n \geq 0$. In case of ties, if the impulse response of interest is to the j-th structural shock, one should order the j-th variable first. If there are only sign restrictions, one should order first the variable whose structural shock is of interest.

The subspace $\mathscr{Q}(\boldsymbol{\phi})$ satisfying the restrictions is then given by

$$\mathscr{Q}(\boldsymbol{\phi}) = \left\{\mathbf{Q} \in \mathscr{O}(n) : F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{\sum_{i=1}^{n} f_i \times 1}, \operatorname{diag}\left(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\right) \geq \mathbf{0}_{n\times 1}\right\}. \tag{4.41}$$

4.5.2.2 Exogeneity restrictions in Proxy SVARs

Proxy SVARs rely on the assumption that there are instruments ('proxies') external to the SVAR that are correlated with particular structural shocks ('relevant') and uncorrelated with other shocks ('exogenous').[4] Set-identification arises in these models when there are multiple proxies for multiple shocks. Giacomini, Kitagawa and Read (2022) propose a robust Bayesian approach to inference in this context that starts by writing the restrictions arising from exogeneity of the proxies as linear constraints on the columns of $\mathbf{Q}$, as follows.

Let $\boldsymbol{\varepsilon}_{(i:j),t} = (\boldsymbol{\varepsilon}_{i,t}, \boldsymbol{\varepsilon}_{i+1,t}, ..., \boldsymbol{\varepsilon}_{j-1,t}, \boldsymbol{\varepsilon}_{j,t})'$ for $i < j$. Assume that $\mathbf{m}_t$ is a $k \times 1$ vector of proxies (with $k < n$) that are correlated with the last k structural shocks, so $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(n-k+1:n),t}) = \boldsymbol{\Psi}$ with $\boldsymbol{\Psi}$ a full-rank matrix, and uncorrelated with the first $n-k$ structural shocks, so $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(1:n-k),t}) = \mathbf{0}_{k\times(n-k)}$. We assume that $\mathbf{m}_t$ follows an SVAR(p_m) with $\boldsymbol{\varepsilon}_t$ included as exogenous variables:

$$\boldsymbol{\Gamma}_0 \mathbf{m}_t = \boldsymbol{\gamma} + \boldsymbol{\Lambda}\boldsymbol{\varepsilon}_t + \sum_{l=1}^{p_m} \boldsymbol{\Gamma}_l \mathbf{m}_{t-l} + \boldsymbol{\nu}_t, \quad t = 1, ..., T, \tag{4.42}$$

with $\boldsymbol{\Gamma}_0$ invertible and $(\boldsymbol{\varepsilon}'_t, \boldsymbol{\nu}'_t)' | \mathscr{F}_{t-1} \sim N(0, \mathbf{I}_{n+k})$, where $\mathscr{F}_{t-1}$ is the information set at time $t-1$. Consider the 'first-stage regression'

$$\mathbf{m}_t = \mathbf{g} + \mathbf{D}\mathbf{y}_t + \mathbf{G}\mathbf{x}_t + \sum_{l=1}^{p_m} \mathbf{H}_l \mathbf{m}_{t-l} + \mathbf{v}_t, \tag{4.43}$$

where $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$ and $\mathbb{E}(\mathbf{v}_t \mathbf{v}'_t) = \boldsymbol{\Upsilon}$. Since $\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda} = \mathbf{D}\mathbf{A}_0^{-1} = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q}$, the instrument validity conditions imply that

$$\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_t) = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q} = \left[\mathbf{0}_{k\times(n-k)}, \boldsymbol{\Psi}\right]. \tag{4.44}$$

---

[4]See, for example, Mertens and Ravn (2013) and Stock and Watson (2018).

The (i, j)th element of this matrix is $\mathbf{e}'_{i,k}\mathbf{D}\mathbf{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_{j,n} = \mathbf{d}'_i\mathbf{q}_j$, where $\mathbf{d}'_i \equiv \mathbf{e}'_{i,k}\mathbf{D}\mathbf{\Sigma}_{tr}$ is the i-th row of $\mathbf{D}\mathbf{\Sigma}_{tr}$. The exogeneity conditions in a proxy SVAR therefore generate linear zero restrictions on the first $n-k$ columns of $\mathbf{Q}$ given $\mathbf{D}$ and $\mathbf{\Sigma}_{tr}$. Similarly to the previous subsection, we can write these restrictions in the general form $F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{k(n-k)\times 1}$, where the reduced-form parameter now additionally contains the reduced-form parameters from the first-stage regression. The subspace $\mathscr{Q}(\boldsymbol{\phi})$ satisfying the restrictions is then defined as in (4.41).

4.5.2.3 Sign restrictions

Sign restrictions on impulse-responses can also be written as linear constraints on the columns of $\mathbf{Q}$: $S_{hj}(\boldsymbol{\phi})\mathbf{q}_j \geq \mathbf{0}_{s_{jh}\times 1}$,[5] where $S_{hj}(\boldsymbol{\phi}) \equiv \mathbf{D}_{hj}\mathbf{C}_h\mathbf{\Sigma}_{tr}$, with $\mathbf{D}_{hj}$ a matrix selecting the sign-restricted responses from $\mathbf{C}_h\mathbf{\Sigma}_{tr}\mathbf{q}_j$, with nonzero element 1 or -1 depending on whether the responses are positive or negative. By stacking $S_{hj}(\boldsymbol{\phi})$ over multiple horizons we obtain the set of sign restrictions on the responses to the j-th shock, $S_j(\boldsymbol{\phi})\mathbf{q}_j \geq \mathbf{0}_{s_j\times 1}$. We represent a collection of sign restrictions, $\{S_j(\boldsymbol{\phi})\mathbf{q}_j \geq \mathbf{0}_{s_j\times 1} \text{ for } j \in \mathscr{I}_S\}$ as

$$S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s\times 1}, \tag{4.45}$$

where $\mathscr{I}_S \subset \{1, 2, \ldots, n\}$ is such that $j \in \mathscr{I}_S$ if some responses to the j-th shock are restricted.

The subspace $\mathscr{Q}(\boldsymbol{\phi})$ satisfying the restrictions is then given by

$$\mathscr{Q}(\boldsymbol{\phi}) = \left\{\mathbf{Q} \in \mathscr{O}(n) : S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s\times 1}, \operatorname{diag}\left(\mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}\right) \geq \mathbf{0}_{n\times 1}\right\}. \tag{4.46}$$

Sign restrictions on other parameters, such as elements of $\mathbf{A}_0$, $\mathbf{CIR}^{\infty}$, or (in proxy SVARs) $\mathbf{\Psi}$, can be imposed similarly.

4.5.3 Multiple priors in SVARs

In a standard Bayesian approach to SVAR estimation, one typically specifies a prior for $\boldsymbol{\theta} = (\boldsymbol{\phi}, \mathbf{Q})$, $\pi_{\boldsymbol{\theta}} = \pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_{\boldsymbol{\phi}}$, by specifying single priors $\pi_{\boldsymbol{\phi}}$ and $\pi_{\mathbf{Q}|\boldsymbol{\phi}}$. As discussed

[5]For $\mathbf{y} = (y_1, \ldots, y_m)'$, $\mathbf{y} \geq 0$ means $y_i \geq 0$ for all $i = 1, 2, \ldots, m$.

in Section 4.3, the former is updated by the data, while the latter is not updated (see also Baumeister and Hamilton (2015)).

Maintaining the choice of a single prior for $\boldsymbol{\phi}$, we discuss three approaches relying on different specifications for the set of priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$: the full ambiguity approach of GK, which applies the set of priors of Section 4.3.3; the model-averaging approach of GKV, which can be viewed as robust Bayesian analysis with an ε-contaminated set (Example 4.2.2); and the robust control approach of GKU, which introduces the KL-neighborhood set of conditional priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$.

4.5.3.1 Full ambiguity (GK)

In terms of the current notation for SVAR applications, the set of conditional priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$ representing full ambiguity for the unrevisable component of the prior can be represented as

$$\Pi_{\boldsymbol{\theta}}^{FA} = \{\pi_{\boldsymbol{\theta}} = \int \pi_{\mathbf{Q}|\boldsymbol{\phi}} d\pi_{\boldsymbol{\phi}}(\boldsymbol{\phi}) : \pi_{\mathbf{Q}|\boldsymbol{\phi}} \in \Pi_{\mathbf{Q}|\boldsymbol{\phi}}\}, \tag{4.47}$$

where $\Pi_{\mathbf{Q}|\boldsymbol{\phi}}$ allows all conditional priors supported on the subspace $\mathscr{Q}(\boldsymbol{\phi})$ of orthonormal matrices,

$$\Pi_{\mathbf{Q}|\boldsymbol{\phi}}^{FA} = \left\{\pi_{\mathbf{Q}|\boldsymbol{\phi}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}}\left(\mathscr{Q}(\boldsymbol{\phi})\right) = 1,\, \pi_{\boldsymbol{\phi}}\text{-almost surely}\right\}. \tag{4.48}$$

As shown in (4.22), the resulting set of posteriors for the impulse response of interest $\eta = h(\mathbf{Q}, \boldsymbol{\phi})$ is

$$\Pi_{\eta|\mathbf{Y}}^{FA} = \left\{\pi_{\eta|\mathbf{Y}}(\cdot) = \int \pi_{\mathbf{Q}|\phi}\left(h(\mathbf{Q}, \boldsymbol{\phi}) \in \cdot\right) d\pi_{\boldsymbol{\phi}|\mathbf{Y}} : \pi_{\mathbf{Q}|\phi} \in \Pi_{\mathbf{Q}|\phi}^{FA}\right\}. \tag{4.49}$$

The general formulae for the set of posterior means (4.24) and the robust credible regions (4.25) shown in Section 4.3.3 can apply as they are. See Section 4.6.1 for algorithms to compute these quantities.

4.5.3.2 Model averaging (GKV)

GKV consider a set of priors for the impulse-response that averages single-prior and multiple-prior models. Focusing on the case of two models, the approach can be viewed as a refinement of GK when the researcher has access to a single prior for the impulse-response; for example, that implied by a prior on the SVAR's structural parameters (e.g., Baumeister and Hamilton (2015)) or a prior based on a Bayesian DSGE model. Another example is when identifying restrictions yield point-identification but some of the restrictions are controversial. The single prior in this case corresponds to the point-identified model imposing all restrictions, while the multiple-prior model corresponds to a set-identified model that relaxes the controversial restrictions.

Let M^p be the single-prior model and M^s the multiple-prior (set-identified) model, with corresponding prior probabilities $\pi_{M^p} \in [0,1]$ and $\pi_{M^s} = 1 - \pi_{M^p}$. The single-prior model admits a unique prior for $\boldsymbol{\theta}$, $\pi_{\boldsymbol{\theta}|M^p}$, while the input of the multiple-prior model is the GK set of priors for $\boldsymbol{\theta}$, $\Pi_{\boldsymbol{\theta}|M^s}$, given a unique prior for $\boldsymbol{\phi}$, $\pi_{\boldsymbol{\phi}|M^s}$. GKV obtain a set of posteriors for the impulse-response that combines the single posterior in model M^p and the set of posteriors in model M^s according to the posterior model probabilities.

This practice of averaging the single-prior (or point-identified) model and the multiple-prior (set-identified) model can be viewed as a robust Bayesian analysis with the following set of priors:

$$\Pi_{\boldsymbol{\theta}}^{Avg} \equiv \left\{ \pi_{\boldsymbol{\theta}} = \pi_{\boldsymbol{\theta}|M^p}\pi_{M^p} + \pi_{\boldsymbol{\theta}|M^s}\pi_{M^s} : \pi_{\boldsymbol{\theta}|M^s} \in \Pi_{\boldsymbol{\theta}|M^s}^{FA} \right\}. \tag{4.50}$$

This set of priors takes the form of an ε-contaminated set of priors as in (4.13), where the benchmark prior is from the single-prior (point-identified) model $\pi_{\boldsymbol{\theta}}^0 = \pi_{\boldsymbol{\theta}|M^p}$, the amount of contamination is the prior model probability assigned to the set-identified model $\varepsilon = \pi_{M^s}$ and $\mathscr{Q}_{\boldsymbol{\theta}}$ corresponds to the full-ambiguity set of priors for the set-identified model $\Pi_{\boldsymbol{\theta}|M^s}^{FA}$. That is, if the single-prior (or point-identified) model is a possibly misspecified benchmark, averaging it with the set-identified model can be interpreted as performing Bayesian sensitivity analysis with respect to a contamination of the benchmark model by an amount π_{M^s} in every possible direction, while maintaining the set-identifying restrictions in M^s.

GKV show that the posterior model probabilities differ from the prior model probabilities if the models are 'distinguishable' for some values of $\boldsymbol{\phi}$ and/or the two models consider different priors for $\boldsymbol{\phi}$. Models are distinguishable if they imply different reduced-form parameter spaces. Models admitting the same reduced-form representation (i.e., a VAR with the same variables and lag length) but differing in the identifying restrictions they impose are distinguishable if the restrictions rule out different values of $\boldsymbol{\phi}$ (e.g., by yielding an empty identified set for some values of $\boldsymbol{\phi}$).

The posterior model probabilities are obtained as

$$\begin{aligned} \pi_{M^p|\mathbf{Y}} &= \frac{p(\mathbf{Y}|M^p)\cdot\pi_{M^p}}{p(\mathbf{Y}|M^p)\cdot\pi_{M^p}+p(\mathbf{Y}|M^s)\cdot\pi_{M^s}}, \\ \pi_{M^s|\mathbf{Y}} &= \frac{p(\mathbf{Y}|M^s)\cdot\pi_{M^s}}{p(\mathbf{Y}|M^p)\cdot\pi_{M^p}+p(\mathbf{Y}|M^s)\cdot\pi_{M^s}}, \end{aligned} \tag{4.51}$$

where $p(\mathbf{Y}|M) \equiv \int p(\mathbf{Y}|\boldsymbol{\phi},M)d\pi_{\boldsymbol{\phi}|M}(\boldsymbol{\phi})$ is the marginal likelihood of model M with $p(\mathbf{Y}|\boldsymbol{\phi},M)$ the likelihood of the reduced-form parameter.

The set of posteriors can again be summarized by reporting a set of posterior means and a robust credible region. The set of posterior means for η is the weighted average of the posterior mean in model M^p and the set of posterior means in model M^s:

$$\begin{aligned} &\left[\inf_{\pi_{\eta|\mathbf{Y}}\in\Pi_{\eta|\mathbf{Y}}} E_{\eta|\mathbf{Y}}(\eta),\ \sup_{\pi_{\eta|\mathbf{Y}}\in\Pi_{\eta|\mathbf{Y}}} E_{\eta|\mathbf{Y}}(\eta)\right] \\ =&\pi_{M^p|\mathbf{Y}}E_{\eta|M^p,\mathbf{Y}}(\eta)+\pi_{M^s|\mathbf{Y}}\left[E_{\boldsymbol{\phi}|M^s,\mathbf{Y}}(\ell(\boldsymbol{\phi})),E_{\boldsymbol{\phi}|M^s,\mathbf{Y}}(u(\boldsymbol{\phi}))\right], \end{aligned} \tag{4.52}$$

where $(\ell(\boldsymbol{\phi}),u(\boldsymbol{\phi}))$ are as defined in (4.24) and $E_{\boldsymbol{\phi}|M^s,\mathbf{Y}}(\cdot)$ is the posterior mean with respect to the $\boldsymbol{\phi}$-prior in model M^s. Section 4.6.2 discusses how to compute the set of posterior moments and the robust credible regions.

A potentially useful analysis that can be carried out in this context is a reverse-engineering exercise that computes the prior weight w one would assign to the restrictions in M^p to obtain a given conclusion. For example, to find the smallest weight such that the set of posterior means is contained in the positive real half-line, one solves for w in the equation

$$\frac{p(\mathbf{Y}|M^p)\cdot w}{p(\mathbf{Y}|M^p)\cdot w + p(\mathbf{Y}|M^s)\cdot(1-w)} E_{\eta|M^p,\mathbf{Y}}(\eta) \\ + \frac{p(\mathbf{Y}|M^s)\cdot(1-w)}{p(\mathbf{Y}|M^p)\cdot w + p(\mathbf{Y}|M^s)\cdot(1-w)} E_{\boldsymbol{\phi}|M^s,\mathbf{Y}}(\ell(\boldsymbol{\phi})) = 0. \quad (4.53)$$

4.5.3.3 KL-neighborhood (GKU)

GKU consider a refinement of the set of priors in GK. The starting point is the availability of a benchmark conditional prior for $\boldsymbol{\theta}$ given $\boldsymbol{\phi}$, $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$. By considering the set of priors in a KL neighborhood of the benchmark prior with a given radius $\lambda > 0$,

$$\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda) = \left\{ \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} : \int_{\Theta} \ln\left(\frac{d\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}}{d\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}}\right) d\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\boldsymbol{\theta}) \leq \lambda,\ \pi_{\boldsymbol{\phi}}\text{–almost surely} \right\}, \quad (4.54)$$

one can obtain a set of posteriors for the impulse-response η as

$$\Pi^{KL}_{\eta|\mathbf{Y}}(\lambda) = \left\{ \pi_{\eta|\mathbf{Y}}(\cdot) = \int \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}(h(\boldsymbol{\theta}) \in \cdot) d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}) : \pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda) \right\}. \quad (4.55)$$

$\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ refines $\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ in the following aspects. First, by choosing a partially credible benchmark prior $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$, one can anchor the set of priors to the plausible one, disregarding from $\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ those that are far from the benchmark prior. Second, any prior $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ is absolutely continuous with respect to the benchmark prior. Hence the support of $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ is contained in that of $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$, and they share a common dominating measure, implying that one can constrain the support of $\pi_{\boldsymbol{\theta}|\boldsymbol{\phi}} \in \Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ by the choice of $\pi^0_{\boldsymbol{\theta}|\boldsymbol{\phi}}$. Third, the choice of $\lambda > 0$ conveniently controls the size of the prior set. Specifically, in terms of the set of posterior means spanned, varying λ from 0 to ∞ lets $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ vary from a single-prior Bayes approach under the benchmark prior to the multiple-prior Bayes approach under $\Pi^{FA}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$. GKU suggest eliciting λ by assessing the set of prior means of η or other parameters that $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ spans and matching it with the researcher's partial prior knowledge.

The set of posteriors obtained in (4.55) can be used for sensitivity analysis by reporting, for example, the set of posterior means of a function of interest $f(\eta)$ (e.g., $f(\eta) = \eta$ or $f(\eta) = 1\{\eta \in D\}$). GKU show that this set of posterior means is given by

$$\left[\int_{\Phi}\left(\int_{-\infty}^{\infty} f(\eta)d\pi^{\ell}_{\eta|\boldsymbol{\phi}}(\eta)\right)d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \int_{\Phi}\left(\int_{-\infty}^{\infty} f(\eta)d\pi^{u}_{\eta|\boldsymbol{\phi}}(\eta)\right)d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi})\right], \tag{4.56}$$

where $\pi^{\ell}_{\eta|\boldsymbol{\phi}}$ and $\pi^{u}_{\eta|\boldsymbol{\phi}}$ are obtained by exponential tilting of the benchmark priors,

$$\begin{aligned}
d\pi^{\ell}_{\eta|\boldsymbol{\phi}} &\equiv \frac{\exp\{-f(\eta)/\kappa^{\ell}_{\lambda}(\boldsymbol{\phi})\}}{\int \exp\{-f(\eta)/\kappa^{\ell}_{\lambda}(\phi)\}d\pi^{0}_{\eta|\boldsymbol{\phi}}} \cdot d\pi^{0}_{\eta|\boldsymbol{\phi}}, \qquad (4.57)\\
d\pi^{u}_{\eta|\boldsymbol{\phi}} &\equiv \frac{\exp\{f(\eta)/\kappa^{u}_{\lambda}(\boldsymbol{\phi})\}}{\int \exp\{f(\eta)/\kappa^{u}_{\lambda}(\boldsymbol{\phi})\}d\pi^{0}_{\eta|\boldsymbol{\phi}}} \cdot d\pi^{0}_{\eta|\boldsymbol{\phi}},\\
\kappa^{\ell}_{\lambda}(\boldsymbol{\phi}) &\equiv \arg\min_{\kappa\geq 0}\left\{\kappa \ln \int \exp\left\{\frac{-f(\eta)}{\kappa}\right\} d\pi^{0}_{\eta|\boldsymbol{\phi}}(\eta) + \kappa\lambda\right\},\\
\kappa^{u}_{\lambda}(\boldsymbol{\phi}) &\equiv \arg\min_{\kappa\geq 0}\left\{\kappa \ln \int \exp\left\{\frac{f(\eta)}{\kappa}\right\} d\pi^{0}_{\eta|\boldsymbol{\phi}}(\eta) + \kappa\lambda\right\},
\end{aligned}$$

where $\pi^{0}_{\eta|\boldsymbol{\phi}}$ is the benchmark conditional prior for η given $\boldsymbol{\phi}$ obtained by marginalizing $\pi^{0}_{\boldsymbol{\theta}|\boldsymbol{\phi}}$ to η. See Section 4.6.3 for how to compute these bounds.

The posterior mean upper bound obtained in (4.56) is also useful for solving the posterior Gamma minimax problem. For instance, let $\delta(\mathbf{Y})$ be an estimator for η and $L(\delta(\mathbf{Y}), \eta)$ be an estimation loss function. The posterior Gamma minimax estimator $\delta_{\lambda}(\mathbf{Y})$ with prior set $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ can be obtained by

$$\delta_{\lambda}(\mathbf{Y}) \in \arg\min_{a} \int_{\Phi}\left[\int_{IS_{\eta}(\boldsymbol{\phi})} L(a,\eta)d\pi^{u}_{\eta|\boldsymbol{\phi}}(\eta)\right] d\pi_{\boldsymbol{\phi}|\mathbf{Y}}(\boldsymbol{\phi}), \tag{4.58}$$

where

$$d\pi^{u}_{\eta|\boldsymbol{\phi}} = \frac{\exp\{L(a,\eta)/\kappa_{\lambda}(a,\boldsymbol{\phi})\}}{\int_{IS_{\eta}(\boldsymbol{\phi})} \exp\{L(a,\eta)/\kappa_{\lambda}(a,\boldsymbol{\phi})\}\, d\pi^{0}_{\eta|\boldsymbol{\phi}}} \cdot d\pi^{0}_{\eta|\boldsymbol{\phi}}$$

and $\kappa_{\lambda}(a, \boldsymbol{\phi}) > 0$ is the unique solution to the following convex minimization:

$$\min_{\kappa\geq 0}\left\{\kappa \ln \int_{IS_{\eta}(\boldsymbol{\phi})} \exp\left\{\frac{L(a,\eta)}{\kappa}\right\} d\pi^{0}_{\eta|\boldsymbol{\phi}} + \kappa\lambda\right\}.$$

We emphasize that $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$ constructed above is distinct from the KL-neighborhood set for the unconditional prior discussed in Example 4.2.5. The main difference is that the set of priors in (4.16) allows multiple priors not only for the conditional prior of $\boldsymbol{\theta}|\boldsymbol{\phi}$ but also for the marginal prior of $\boldsymbol{\phi}$. Having multiple priors for $\boldsymbol{\phi}$ enables one to assess posterior sensitivity with respect to the prior for the identifiable reduced-form parameter, but masks the shape of the posterior distributions if the set contains a prior that fits the data poorly. This is because obtaining a large set of posteriors could be due to some priors for $\boldsymbol{\phi}$ that are severely in conflict with the data, rather than indicating a lack of information in the observed likelihood. With $\Pi^{KL}_{\boldsymbol{\theta}|\boldsymbol{\phi}}(\lambda)$, in contrast, all the posteriors in the set share the same value of the marginal likelihood, so we can assess posterior sensitivity while keeping the denominator of Bayes' rule constant.

4.5.4 Frequentist properties

The methods described in the previous sections are valid from a Bayesian perspective as tools for robust Bayesian sensitivity analysis, but it can also be important to understand their frequentist properties. This section briefly summarizes the asymptotic frequentist properties of the multiple-prior estimation and inference procedures covered in this paper, and overviews other approaches. See the individual papers for precise regularity conditions, formal statements of the frequentist results, and proofs.

As shown by Chen et al. (2018), when joint inference for the whole set of structural parameters $\boldsymbol{\theta}$ is concerned, the Bayesian highest posterior density regions under a single prior for $\boldsymbol{\theta}$ asymptotically attain correct frequentist coverage even when $\boldsymbol{\theta}$ is set-identified. Such asymptotic agreement breaks down if we consider a parameter of interest $\eta = h(\boldsymbol{\theta})$ that is of lower dimension than θ.[6] In the setting considered in Section 4.3, GK examine whether the robust Bayesian approach re-

[6]More precisely, asymptotic frequentist validity of the Bayesian highest posterior density regions holds if the identified set for η is a lower-dimensional subregion of the parameter space of η; see Remark 3 in Moon and Schorfheide (2012) and Remark 4 in Chen et al. (2018). In typical parametric set-identified models (including SVARs), the identified set for θ is a lower-dimensional subregion of Θ so that standard Bayesian inference attains frequentist coverage asymptotically, while for a scalar parameter of interest (e.g., one impulse response) with identified set of positive width, the asymptotic undercoverage result of Moon and Schorfheide (2012) applies.

stores the asymptotic equivalence between frequentist and Bayesian inference for η. Under the assumptions that the Bernstein-von Mises property holds for estimation of the reduced-form parameter and that the identified-set mapping for η is convex, continuous and differentiable, the set of posterior means is consistent and the robust credible region has valid frequentist coverage for the true identified set asymptotically. GK also provide primitive conditions in SVARs under which these conditions are satisfied; these conditions typically require checking the pattern of zero and sign restrictions imposed. See also Liao and Simoni (2013) and Kline and Tamer (2016) for Bernstein-von Mises results about Bayesian confidence sets for the identified set.

Chen et al. (2018) develop an inference procedure for the identified set for η that makes use of Monte Carlo draws of $\boldsymbol{\theta}$ from its joint posterior and that relies on profiling the likelihood. By establishing a Bernstein-von Mises property for the posterior of the quasi-likelihood ratio criterion, they show asymptotic frequentist validity of their procedure. In applications where the approach is computationally feasible, the procedure has the advantage of attaining asymptotic frequentist validity without requiring differentiability of the identified-set mapping.

Giacomini, Kitagawa and Read (2022) provide conditions for asymptotic frequentist validity of the robust Bayesian approach in the case of proxy SVARs, and discuss how the case of proxies that are only weakly correlated with the structural shocks ('weak proxies') affects the asymptotic frequentist properties of the GK robust Bayesian credible sets.

The asymptotic frequentist validity of robust Bayesian inference shown in GK reveals that, for set-identified models, ambiguity about the parameters represented by $\Pi_{\boldsymbol{\theta}}^{FA}$ can match the absence or removal of a prior in frequentist inference. This means that robust Bayesian inference under a set of priors Π_θ that is a strict subset of $\Pi_{\boldsymbol{\theta}}^{FA}$ generally leads to more informative inference than frequentist inference when the models are set-identified. Accordingly, the robust credible regions obtained under the GKV set of priors $\Pi_{\boldsymbol{\theta}}^{Avg}$ and the GKU set of priors $\Pi_{\boldsymbol{\theta}}^{KL}(\lambda)$, $\lambda < \infty$, yield posterior inference that is too optimistic in terms of frequentist coverage.

4.6 Numerical Implementation

This section explains how to numerically implement the three approaches described in the previous section. We emphasise the key choices practitioners face when implementing the algorithms.

4.6.1 Full ambiguity (GK)

We present a general algorithm to numerically approximate the set of posterior means and the robust credible region. The algorithm assumes $f_i \leq n-i$ for all $i = 1, \ldots, n$.

Algorithm 4.6.1. ***Robust Bayesian inference under full ambiguity (GK).*** *Let* $F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{\sum_{i=1}^{n} f_i \times 1}$ *and* $S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s \times 1}$ *be the set of identifying restrictions, and let* $\eta = \mathbf{c}'_{ih}(\boldsymbol{\phi})\mathbf{q}_{j^*}$ *be the impulse response of interest.*

- ***Step 1:*** *Specify a prior for the reduced-form parameter,* $\tilde{\pi}_{\boldsymbol{\phi}}$.

- ***Step 2:*** *Draw* $\boldsymbol{\phi}$ *from its posterior,* $\tilde{\pi}_{\boldsymbol{\phi}|\mathbf{Y}}$, *and check whether the set of orthonormal matrices satisfying the identifying restrictions,* $\mathcal{Q}(\boldsymbol{\phi})$, *is empty. If so, repeat Step 2. Otherwise, proceed to Step 3.*

- ***Step 3:*** *Given* $\boldsymbol{\phi}$ *obtained in Step 2, compute the lower bound,* $\ell(\boldsymbol{\phi})$, *and upper bound,* $u(\boldsymbol{\phi})$, *of the identified set for* η, $IS_\eta(\boldsymbol{\phi})$. $\ell(\boldsymbol{\phi})$ *is defined by the following minimization problem:*

$$
\begin{aligned}
& \ell(\boldsymbol{\phi}) = \arg\min_{\mathbf{Q}} \mathbf{c}'_{ih}(\boldsymbol{\phi})\mathbf{q}^*_j, \\
\textit{s.t.} \quad & \mathbf{Q}'\mathbf{Q} = \mathbf{I}_n,\ F(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{0}_{\sum_{i=1}^{n} f_i \times 1}, \quad \operatorname{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n \times 1}, \quad S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s \times 1},
\end{aligned}
$$

  and $u(\boldsymbol{\phi}) = \arg\max_{\mathbf{Q}} \mathbf{c}'_{ih}(\boldsymbol{\phi})\mathbf{q}^*_j$ *under the same set of constraints.*

- ***Step 4:*** *Repeat Steps 2–3 M times to obtain* $[l(\boldsymbol{\phi}_m), u(\boldsymbol{\phi}_m)]$, $m = 1, \ldots, M$. *Approximate the set of posterior means by the sample averages of* $l(\boldsymbol{\phi}_m)$ *and* $u(\boldsymbol{\phi}_m)$.

- ***Step 5:*** *To obtain an approximation of the smallest robust credible region with credibility* $\alpha \in (0,1)$, *define* $d(\eta, \phi) = \max\{|\eta - \ell(\boldsymbol{\phi})|, |\eta - u(\boldsymbol{\phi})|\}$, *and let* $\hat{z}_\alpha(\eta)$ *be the sample* α*-th quantile of* $(d(\eta,\boldsymbol{\phi}_m): m=1,\ldots,M)$*. An approximated smallest robust credible region for* η *is an interval centered at* $\arg\min_\eta \hat{z}_\alpha(\eta)$ *with radius* $\min_\eta \hat{z}_\alpha(\eta)$.[7]

The method used to draw $\boldsymbol{\phi}$ from its posterior in Step 2 depends on the prior specified in Step 1 (which may be improper). A commonly used prior for $\boldsymbol{\phi}$ is the normal-inverse-Wishart (e.g., Arias et al. (2018)); this prior induces a normal-inverse-Wishart posterior, from which it is easy to obtain independent draws (e.g., Del Negro and Schorfheide (2011)). It is also possible to apply this algorithm when the prior is specified for the structural parameters rather than the reduced-form parameters, provided the prior for the structural parameters embeds exact zero restrictions and/or dogmatic sign restrictions (e.g., Baumeister and Hamilton (2015)). In this case, draws of the structural parameters (e.g., obtained via Markov Chain Monte Carlo methods) can be transformed into draws of the reduced-form parameters. When the identified set is empty at some values of $\boldsymbol{\phi}$ receiving positive prior probability under $\tilde{\pi}_{\boldsymbol{\phi}}$, the prior for $\boldsymbol{\phi}$ is implicitly trimmed by the algorithm to support only values of $\boldsymbol{\phi}$ yielding a nonempty identified set.

Step 2 requires checking whether the identified set for $\mathbf{Q}$ given $\boldsymbol{\phi}$ is empty. In the case where there are zero restrictions only subject to $f_i \leq n-i$, $i=1,\ldots,n$, the identified set is never empty. When there are sign restrictions (possibly alongside zero restrictions), there are different algorithms to check whether the identified set is empty, and their applicability depends on the types of restriction. The following algorithm in GK can be applied to any pattern of zero and sign restrictions.

Algorithm 4.6.2. ***Checking emptiness of identified set via rejection sampling.***

- ***Step 1:*** *Draw* $\mathbf{z}_1 \sim N(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$ *and let* $\tilde{\mathbf{q}}_1 = \left[\mathbf{I}_n - \mathbf{F}_1'(\mathbf{F}_1\mathbf{F}_1')^{-1}\mathbf{F}_1\right]\mathbf{z}_1$*, then, for* $i=2,\ldots,n$*, draw* $\mathbf{z}_i \sim N(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$ *and compute* $\tilde{\mathbf{q}}_i = \left[\mathbf{I}_n - \tilde{\mathbf{F}}_i'(\tilde{\mathbf{F}}_i\tilde{\mathbf{F}}_i')^{-1}\tilde{\mathbf{F}}_i\right]\mathbf{z}_i$*, where* $\mathbf{F}_i \equiv F_i(\phi)$ *and* $\tilde{\mathbf{F}}_i' = [\mathbf{F}_i', \tilde{\mathbf{q}}_1, \ldots, \tilde{\mathbf{q}}_{i-1}]$.

[7] The objective function in this minimization is nondifferentiable in η, so we recommend obtaining this interval via grid search.

- ***Step 2:*** *Compute*

$$\mathbf{Q}_0 = \left[\text{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\tilde{\mathbf{q}}_1)\frac{\tilde{\mathbf{q}}_1}{\|\tilde{\mathbf{q}}_1\|}, \ldots, \text{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n,n})'\tilde{\mathbf{q}}_n)\frac{\tilde{\mathbf{q}}_n}{\|\tilde{\mathbf{q}}_n\|}\right].$$

- ***Step 3:*** *Check whether $\mathbf{Q}_0$ satisfies $S(\boldsymbol{\phi}, \mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$. If so, conclude $\mathcal{Q}(\boldsymbol{\phi})$ is nonempty. Otherwise, repeat Steps 1–2 (up to a maximum of L times) until $\mathbf{Q}_0$ is obtained satisfying $S(\boldsymbol{\phi}, \mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$. If no draws of $\mathbf{Q}_0$ satisfy $S(\boldsymbol{\phi}, \mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$, approximate $\mathcal{Q}(\boldsymbol{\phi})$ as being empty.*

Step 1 of this algorithm generates orthogonal vectors $(\tilde{\mathbf{q}}_1, ..., \tilde{\mathbf{q}}_n)$ satisfying the zero restrictions. Step 2 normalises these vectors to have unit length and imposes the sign normalization that the diagonal elements of $\mathbf{A}_0$ are nonnegative. The resulting $\mathbf{Q}_0$ is an orthonormal matrix satisfying the zero restrictions and the sign normalizations. Step 3 checks whether the drawn $\mathbf{Q}_0$ satisfies the sign restrictions. The advantage of this algorithm is its generality. A drawback is that, for a finite value of L, the algorithm may misclassify the identified set as being empty. Increasing the value of L reduces the chance of this happening, but at the cost of increased computing time when the identified set is actually empty at some values of $\boldsymbol{\phi}$. In practice, practitioners using this algorithm should check whether the chosen L is large enough by seeing whether the proportion of draws with empty identified set is sensitive to an increase in L.

When there are zero and sign restrictions on a single column of $\mathbf{Q}$, emptiness of the identified set can be determined without recourse to random sampling by using the following algorithm in GKV:

Algorithm 4.6.3. ***Checking emptiness of identified set via active-set algorithm.*** *Assume any zero and sign restrictions apply to $\mathbf{q}_1$ only and let the value of $\boldsymbol{\phi}$ be given. Let $S(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$ represent the sign restrictions (including the sign normalization). Further, assume that the $(n-1)\times n$ matrix $Z(\boldsymbol{\phi}) = [F(\boldsymbol{\phi})', \tilde{S}(\boldsymbol{\phi})']'$ has rank $n-1$ for any $(n-f_1-1)\times n$ matrix $\tilde{S}(\boldsymbol{\phi})$ constructed from a selection of $n-f_1-1$ rows of $S(\boldsymbol{\phi})$.*

- ***Step 1:*** *Choose $n - f_1 - 1$ rows from $S(\boldsymbol{\phi})$ and collect these in $\tilde{S}(\boldsymbol{\phi})$. Construct $Z(\boldsymbol{\phi}) = [F(\boldsymbol{\phi})', S(\boldsymbol{\phi})']'$.*

- ***Step 2:*** *Compute an orthonormal basis for the null space of $Z(\boldsymbol{\phi})$, $N(Z(\boldsymbol{\phi}))$, which is an $n \times 1$ vector.*

- ***Step 3:*** *Check if either $N(Z(\boldsymbol{\phi}))$ or $-N(Z(\boldsymbol{\phi}))$ satisfies the remaining $s - (n - f_1 - 1)$ sign restrictions not contained in $\tilde{S}(\boldsymbol{\phi})$. If so, conclude $\mathscr{Q}(\boldsymbol{\phi})$ is nonempty. Otherwise, return to Step 1 until all $\binom{s}{n-f-1}$ combinations are exhausted, in which case conclude $\mathscr{Q}(\boldsymbol{\phi})$ is empty.*

This algorithm relies on the fact that any nonempty identified set for $\mathbf{q}_1$ must contain a vertex on the unit sphere where at least $n-1$ constraints are binding. The algorithm determines whether the identified set is nonempty by considering all possible combinations of $n - f_1 - 1$ binding sign restrictions and checking whether the implied vertex satisfies the remaining $s - (n - f_1 - 1)$ sign restrictions. Under the assumptions stated in the algorithm, the rank-nullity theorem implies that the null space of $Z(\boldsymbol{\phi})$ is one-dimensional, but if some $\mathbf{q}$ satisfies $Z(\boldsymbol{\phi})\mathbf{q} = \mathbf{0}_{(n-1)\times 1}$, then so too does $-\mathbf{q}$, so it is necessary to check whether $N(Z(\boldsymbol{\phi}))$ or $-N(Z(\boldsymbol{\phi}))$ satisfies the sign restrictions excluded from $\tilde{S}(\boldsymbol{\phi})$. The advantage of this algorithm over Algorithm 4.6.2 is that it will never misclassify the identified set as being nonempty. However, since the algorithm requires checking $\binom{s}{n-f-1}$ combinations of restrictions, the algorithm may become slow or infeasible when there is a large number of sign restrictions.

A third approach to checking whether the identified set is empty is the 'Chebyshev criterion' proposed in Amir-Ahmadi and Drautzburg (2021). This algorithm is applicable when there are sign restrictions constraining a single column of $\mathbf{Q}$ and there are no zero restrictions. Read (in press) extends this algorithm to allow for zero restrictions. These algorithms are useful when there are many sign restrictions that appreciably truncate the identified set.

Step 3 of Algorithm 1 requires computing the bounds of the identified set for η at each draw of $\boldsymbol{\phi}$. If one is interested in more than one scalar object at a time (e.g., impulse responses for multiple variables), this step is run repeatedly at each draw of $\boldsymbol{\phi}$. As for Step 2, there are multiple approaches for computing the bounds of the identified set. GK suggest two different approaches that are applicable under arbitrary configurations of zero and sign restrictions. The first is to use a numerical optimizer initialised at the value of $\mathbf{Q}_0$ obtained using Algorithm 2. This is a nonconvex optimization problem, so convergence to the true optimum is not guaranteed.

The second approach is to repeat Algorithm 2 many times at each draw of $\boldsymbol{\phi}$ to obtain a large number of draws of $\mathbf{Q}$ from $\mathscr{Q}(\boldsymbol{\phi})$ and then to compute the minimum and maximum of η over the draws. This provides an approximated identified set that is smaller than the actual identified set, but that converges to the actual identified set as the number of draws goes to infinity. A third approach is available when the zero and sign restrictions constrain a single column of $\mathbf{Q}$ only, in which case the bounds of the identified set can be computed using the active-set algorithm in Gafarov et al. (2018). This approach may be prohibitively slow when there are many sign restrictions.

4.6.2 Model averaging (GKV)

This section presents a general algorithm to numerically approximate the set of posterior means when there is uncertainty over the set of identifying restrictions, as in GKV.

Algorithm 4.6.4. ***Uncertain identification (GKV).***

- ***Step 1:*** *Draw a model $M \in \mathscr{M}$ from a multinomial distribution with parameters $(\pi_{M|\mathbf{Y}} : M \in \mathscr{M})$.*

- ***Step 2:*** *If the drawn M belongs to $\mathscr{M}_p$, draw $\eta \sim \pi_{\eta|M,\mathbf{Y}}$ and set $IS_\eta^{mix} = \{\eta\}$. If the drawn M belongs to $\mathscr{M}_s$, draw $\boldsymbol{\phi}_M \sim \pi_{\boldsymbol{\phi}|M,\mathbf{Y}}$ and set $IS_\eta^{mix} = IS_\eta(\boldsymbol{\phi}_M|M)$.*

- ***Step 3:*** *Repeat Steps 1 and 2 G times to obtain G draws of IS_η^{mix}.*

- ***Step 4:*** *Let l_g^{mix} and u_g^{mix} be the upper and lower bounds of IS_η^{mix}, respectively, where $l_g^{mix} = u_g^{mix}$ if the gth draw of M belongs to $\mathcal{M}_p$. Approximate the bounds of the set of posterior means by the sample averages of l_g^{mix} and u_g^{mix}.*

Step 1 of this algorithm requires computing the posterior model probabilities, $\pi_{M|\mathbf{Y}}$, for each model or, equivalently, the marginal likelihood for each model. Algorithms to compute the marginal likelihood include those in Chib and Jeliazkov (2001), Geweke (1999) and Sims et al. (2008). When the models admit an identical reduced form, it is unnecessary to compute the marginal likelihoods, because the posterior model probabilities depend only on posterior-prior plausibility ratios, which are the posterior probability that the identified set is nonempty divided by the prior probability that the identified set is nonempty in each model. These ratios can be computed using numerical approximations of the prior and posterior probabilities that the identified set is nonempty. Depending on the pattern of zero and sign restrictions considered, these probabilities can be computed by drawing $\boldsymbol{\phi}$ from its prior or posterior and using the algorithms described in the previous subsection to check whether the identified set is nonempty. Since computing these ratios requires drawing $\boldsymbol{\phi}$ from its prior, it is necessary for this prior to be proper.[8]

Step 2 requires drawing from the posterior of the object of interest η when the model is point-identified or the prior is for the structural parameters, which are standard problems. For example, when the prior is for the reduced-form parameters, one simply draws from the posterior for $\boldsymbol{\phi}$ and transforms the draw. When the sampled model is set-identified, Step 2 requires drawing $\boldsymbol{\phi}$ from its posterior and computing the identified set for the object of interest. Note that $\pi_{\boldsymbol{\phi}|M,\mathbf{Y}}$ supports only values of $\boldsymbol{\phi}$ with nonempty identified set. In practice, the practitioner may specify a prior assigning positive probability to regions of the reduced-form parameter space with empty identified set and simply continue to draw $\boldsymbol{\phi}$ from its posterior at a given draw of M until the identified set is nonempty. The bounds of the identified set

[8]It is possible to eliminate a source of Monte Carlo sampling variability arising from Step 1 by avoiding sampling M from a multinomial distribution. For example, if the posterior model probabilities are known to be 0.5 when averaging over two models, one simply needs to draw $G/2$ times from the relevant posterior for each model.

can be computed using any of the approaches described in the previous subsection (depending on the pattern of zero and sign restrictions). As in GK, the draws of l_g^{mix} and u_g^{mix} can be used to construct a robust credible interval.

4.6.3 KL-neighborhood (GKU)

This section discusses how to compute the set of posterior means and the posterior Gamma minimax decision (i.e., the point-estimator under ambiguity). The algorithms below assume that posterior draws of $\boldsymbol{\phi}$ are given and that η can be drawn from the benchmark conditional prior. GKU also present modifications of these algorithms when direct draws of η are unavailable but its probability density can be evaluated up to a proportional constant. The first algorithm below describes how to compute the set of posterior means for some object of interest $f(\eta)$.

Algorithm 4.6.5. ***Set of posterior means given KL neighborhood.*** *Let G posterior draws of* $\boldsymbol{\phi}$*,* $\{\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_G\}$*, and benchmark conditional prior* $\pi^*_{\eta|\boldsymbol{\phi}}$ *be given.*

- ***Step 1:*** *For each* $g=1,\ldots,G$*, obtain K independent draws of* η*,* $\eta_{gk}\sim\pi^*_{\eta|\boldsymbol{\phi}}$*,* $k=1,\ldots,K$*. Approximate the Lagrange multipliers,* $\kappa^l_\lambda(\boldsymbol{\phi}_g)$ *and* $\kappa^u_\lambda(\boldsymbol{\phi}_g)$*, by solving the following optimization problems:*

$$\hat{\kappa}^l_\lambda(\boldsymbol{\phi}_g)\equiv\arg\min_{\kappa\geq 0}\left\{\kappa\ln\left(\frac{1}{K}\sum_{k=1}^{K}\exp\left(-\frac{f(\eta_{gk})}{\kappa}\right)\right)+\kappa\lambda\right\} \tag{4.59}$$

$$\hat{\kappa}^u_\lambda(\boldsymbol{\phi}_g)\equiv\arg\min_{\kappa\geq 0}\left\{\kappa\ln\left(\frac{1}{K}\sum_{k=1}^{K}\exp\left(\frac{f(\eta_{gk})}{\kappa}\right)\right)+\kappa\lambda\right\}. \tag{4.60}$$

- ***Step 2:*** *Approximate the set of posterior means of* $f(\eta)$ *by*

$$\left[\frac{1}{G}\sum_{g=1}^{G}\left(\frac{\sum_{k=1}^{K}f(\eta_{gk})\exp\left(-\frac{f(\eta_{gk})}{\hat{\kappa}^l_\lambda(\phi_g)}\right)}{\sum_{k=1}^{K}\exp\left(-\frac{f(\eta_{gk})}{\hat{\kappa}^l_\lambda(\phi_g)}\right)}\right),\frac{1}{G}\sum_{g=1}^{G}\left(\frac{\sum_{k=1}^{K}f(\eta_{gk})\exp\left(\frac{f(\eta_{gk})}{\hat{\kappa}^u_\lambda(\phi_g)}\right)}{\sum_{k=1}^{K}\exp\left(\frac{f(\eta_{gk})}{\hat{\kappa}^u_\lambda(\phi_g)}\right)}\right)\right]. \tag{4.61}$$

The optimization problems in Step 1 are convex and can be solved reliably using a gradient-based numerical optimization routine, such as the interior-point algorithm in Matlab's 'fmincon' optimizer. Care should be taken when the Lagrange multipliers are close to zero, since terms in the expression for the set of posterior means may be very large and result in numerical overflow.

The following algorithm can be used to approximate the worst-case risk of decision (i.e., estimator) a, which can then be used to compute the posterior Gamma minimax estimator.

Algorithm 4.6.6. ***Computing worst-case risk.*** *Let G posterior draws of $\boldsymbol{\phi}$, $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_G\}$, and benchmark conditional prior $\pi^*_{\eta|\boldsymbol{\phi}}$ be given. Let $h(a,\eta)$ be the loss function (e.g., quadratic or check).*

- ***Step 1:*** *For each $g = 1, \ldots, G$, obtain K independent draws of η, $\eta_{gk} \sim \pi^*_{\eta|\boldsymbol{\phi}}$, $k = 1, \ldots, K$. Approximate the Lagrange multiplier, $\kappa_\lambda(\boldsymbol{\phi}_g)$, by solving the following optimization problem:*

$$\hat{\kappa}_\lambda(a, \boldsymbol{\phi}_g) \equiv \arg\min_{\kappa \geq 0} \left\{ \kappa \ln \left(\frac{1}{K} \sum_{k=1}^{K} \exp\left(\frac{h(a, \eta_{gk}))}{\kappa} \right) \right) + \kappa\lambda \right\}. \tag{4.62}$$

- ***Step 2:*** *For each $g = 1, \ldots, G$, compute*

$$\hat{r}_\lambda(a, \boldsymbol{\phi}_g) = \frac{\sum_{k=1}^{K} h(a, \eta_{gk}) \exp\left(\frac{h(a,\eta_{gk})}{\hat{\kappa}_\lambda(a,\boldsymbol{\phi}_g)} \right)}{\sum_{k=1}^{K} \exp\left(\frac{h(a,\eta_{gk})}{\hat{\kappa}_\lambda(a,\boldsymbol{\phi}_g)} \right)}. \tag{4.63}$$

The posterior Gamma minimax estimator is then obtained by minimising $(1/G)\sum_{g=1}^{G} \hat{r}_\lambda(a, \boldsymbol{\phi}_g)$ with respect to a. If the loss is differentiable in a, this minimization can be carried out using a gradient-based numerical optimization routine. Otherwise, the minimization can be done via grid search, which is not computationally costly because a is a scalar. In either case, the same draws of $\boldsymbol{\phi}$ and η (at each draw of $\boldsymbol{\phi}$) should be used for the optimization.

4.7 Empirical Illustration

This section illustrates how to apply the methods described above using an empirical example. The empirical application considered is from Arias et al. (2019; henceforth, ACR), who estimate the effects of monetary policy shocks in the United

States using a mixture of zero and sign restrictions on the systematic response of the federal funds rate to macroeconomic variables.

Reduced-form VAR. The model's endogenous variables are real GDP (GDP_t), the GDP deflator ($GDPDEF_t$), a commodity price index (COM_t), total reserves (TR_t), non-borrowed reserves (NBR_t) (all in natural logarithms) and the federal funds rate (FFR_t). The data are monthly and run from January 1965 to June 2007. The VAR includes 12 lags and no deterministic terms.

In order to apply the approach in GKV, we need to compute the posterior-prior plausibility ratio. This requires drawing from the prior for $\boldsymbol{\phi}$, so this prior needs to be proper. ACR use an improper normal-inverse-Wishart prior, which is inappropriate for our purposes. Instead, we use a diffuse (but proper) normal-inverse-Wishart prior under which the prior means of the VAR coefficients imply each variable in $\mathbf{y}_t$ follows a univariate random walk a priori. The posterior for the reduced-form parameters is then also a normal-inverse-Wishart distribution, from which it is straightforward to obtain independent draws (e.g., using the sampler described in Del Negro and Schorfheide (2011)).[9]

Identifying restrictions. Let $\mathbf{y}_t = (FFR_t, GDP_t, GDPDEF_t, COM_t, TR_t, NBR_t)'$. The monetary policy shock is ε_{1t} and the first equation of the SVAR can be interpreted as the monetary policy reaction function. ACR19 set-identify impulse responses to the monetary policy shock using a mixture of sign and zero restrictions on the monetary policy reaction function. The zero restrictions they impose are that FFR_t does not react contemporaneously to TR_t and NBR_t (i.e., these two variables do not appear in the central bank's reaction function), which implies $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{5,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{5,6})'\mathbf{q}_1 = 0$ and $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{6,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{6,6})'\mathbf{q}_1 = 0$. The sign restrictions they impose are that the contemporaneous reactions of FFR_t to GDP_t and $GDPDEF_t$ are nonnegative, which – given the sign normalization $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{1,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,6})'\mathbf{q}_1 \geq 0$ – implies $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{2,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{2,6})'\mathbf{q}_1 \leq 0$ and $\mathbf{e}_{1,6}'\mathbf{A}_0\mathbf{e}_{3,6} = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{3,6})'\mathbf{q}_1 \leq 0$. Additionally, the impact impulse response of the federal funds rate to a monetary policy

[9]The results when conducting standard Bayesian inference (as in ACR) and robust Bayesian inference (as in GK) reported below are very similar to those obtained using the same (improper) prior used in ACR.

shock is restricted to be nonnegative, which implies $\mathbf{e}_{1,6}'\mathbf{A}_0^{-1}\mathbf{e}_{1,6} = \mathbf{e}_{1,6}'\boldsymbol{\Sigma}_{tr}\mathbf{q}_1 \geq 0$.

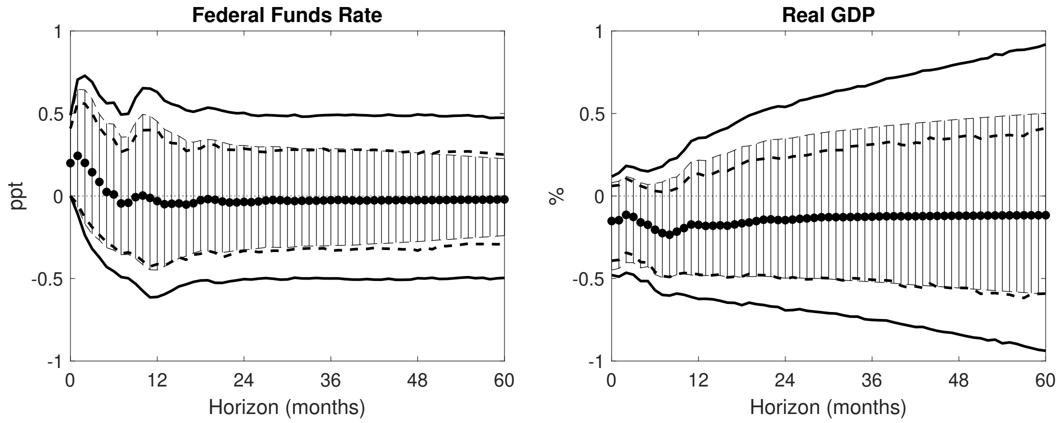
Standard Bayesian inference. The approach to Bayesian inference used in ACR assumes a uniform prior for $\mathbf{Q}$. Here, we assume the conditional prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ is uniform over the space of orthonormal matrices satisfying the identifying restrictions.[10] We obtain 10,000 independent draws from the normal-inverse-Wishart posterior for $\boldsymbol{\phi}$ such that the identified set is nonempty; we check whether the identified set is nonempty using Algorithm 4.6.3. At each draw of $\boldsymbol{\phi}$, we obtain a draw of $\mathbf{Q}$ from the uniform distribution over $\mathscr{Q}(\boldsymbol{\phi}|S)$ using Algorithm 4.6.2. The resulting joint draw of $(\boldsymbol{\phi}, \mathbf{Q})$ is then used to compute the impulse responses.

Figure 4.1 plots the impulse responses of the federal funds rate and real GDP to a positive standard-deviation monetary policy shock (the responses of the remaining variables are omitted for brevity). Based on the posterior mean (black circles), the federal funds rate increases by about 20 basis points in response to the shock before declining to be around its pre-shock value after six months. Output is around 0.2 per cent lower in the year after the shock and the posterior probability of a negative output response is reasonably high at short horizons (e.g., around 85 per cent on impact and one year after the shock). Overall, the results essentially replicate those in ACR and suggest that a positive monetary policy shock results in an economic contraction.

Full ambiguity (GK). As discussed above, the results obtained under the single prior may be sensitive to the choice of prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$, which is not updated by the data. To address this concern, the robust Bayesian approach of GK replaces the unrevisable prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ with the set of all (conditional) priors that are consistent with the identifying restrictions in the sense that the prior places probability one on the identified set given $\boldsymbol{\phi}$. This generates a set of posteriors, which can be summarised by a set of posterior means (an estimator of the identified set) and a

[10]This differs slightly from the prior used in ACR, who assume a uniform-normal-inverse-Wishart prior for $\mathbf{Q}$ and $\boldsymbol{\phi}$. In terms of implementation, our prior requires a single draw of $\mathbf{Q}$ to be obtained at each draw of $\boldsymbol{\phi}$ (with nonempty identified set). In contrast, the prior in ACR is imposed by making a joint draw of $\boldsymbol{\phi}$ from the normal-inverse-Wishart posterior and $\mathbf{Q}$ from the uniform distribution over $\mathscr{O}(n)$, and rejecting joint draws violating the sign restrictions. See Uhlig (2017) for a discussion of this point. This difference in priors does not substantively affect the results.

Figure 4.1: Impulse Responses to a Monetary Policy Shock – Standard and Robust Bayesian Inference



Notes: Circles and dashed lines are, respectively, posterior means and 95 per cent (pointwise) highest posterior density intervals under the uniform prior for $\mathbf{Q}|\boldsymbol{\phi}$. Vertical bars are sets of posterior means and solid lines are 95 per cent (pointwise) robust credible regions. Impulse responses are to a standard-deviation shock.

robust credible region (the shortest interval assigned at least a given posterior probability under all posteriors within the set). To obtain these quantities, it is necessary to compute the lower and upper bound of the identified set for the object of interest (i.e., the impulse response at each horizon) at each draw of $\boldsymbol{\phi}$. As discussed above, there are several ways to this. Here, we apply the active-set algorithm described in Gafarov et al. (2018) at each draw of $\boldsymbol{\phi}$ from its posterior.

The set of posterior means (the vertical lines) includes zero at all horizons. This means there exist (unrevisable) priors for $\mathbf{Q}$ given $\boldsymbol{\phi}$ that are consistent with the identifying restrictions and yield positive posterior mean output responses. The posterior lower probability – the lowest probability over all posteriors generated by the set of priors – of a negative output response is zero at all horizons considered.[11] The result that output falls with high posterior probability obtained under standard Bayesian inference is therefore sensitive to the choice of unrevisable prior. GK propose to quantify the influence of the choice of single prior on posterior inference by comparing the width of the highest posterior density intervals (dashed lines) against that of the robust Bayesian credible intervals (solid lines). On average across the

[11]The posterior lower probability that the output response is negative at a given horizon is approximated by the share of draws from the posterior of $\boldsymbol{\phi}$ where the upper bound of the identified set for the output response is negative (i.e., $u(\boldsymbol{\phi}) < 0$).

horizons considered, the width of the 95 per cent highest posterior density intervals for the output response is 40 per cent that of the robust credible intervals, which suggests that the unrevisable prior contributes a substantial amount of the information contained in the standard Bayesian posterior.

Model averaging (GKV). To illustrate the application of the (robust) Bayesian model-averaging procedure in GKV, we consider a second set of identifying restrictions in addition to the set considered above. Specifically, we use the classic recursiveness assumption considered in, for example, Christiano, Eichenbaum and Evans (1999). Under this set of restrictions, GDP_t, $GDPDEF_t$ and COM_t do not respond contemporaneously to a monetary policy shock (i.e., a shock to FFR_t), while FFR_t does not respond contemporaneously to shocks in TR_t and NBR_t. After re-ordering the variables so $y_t = (GDP_t, GDPDEF_t, COM_t, FFR_t, TR_t, NBR_t)'$, the restrictions imply $\mathbf{A}_0^{-1}$ is block lower-triangular. These restrictions are sufficient to point-identify the impulse responses to a monetary policy shock.

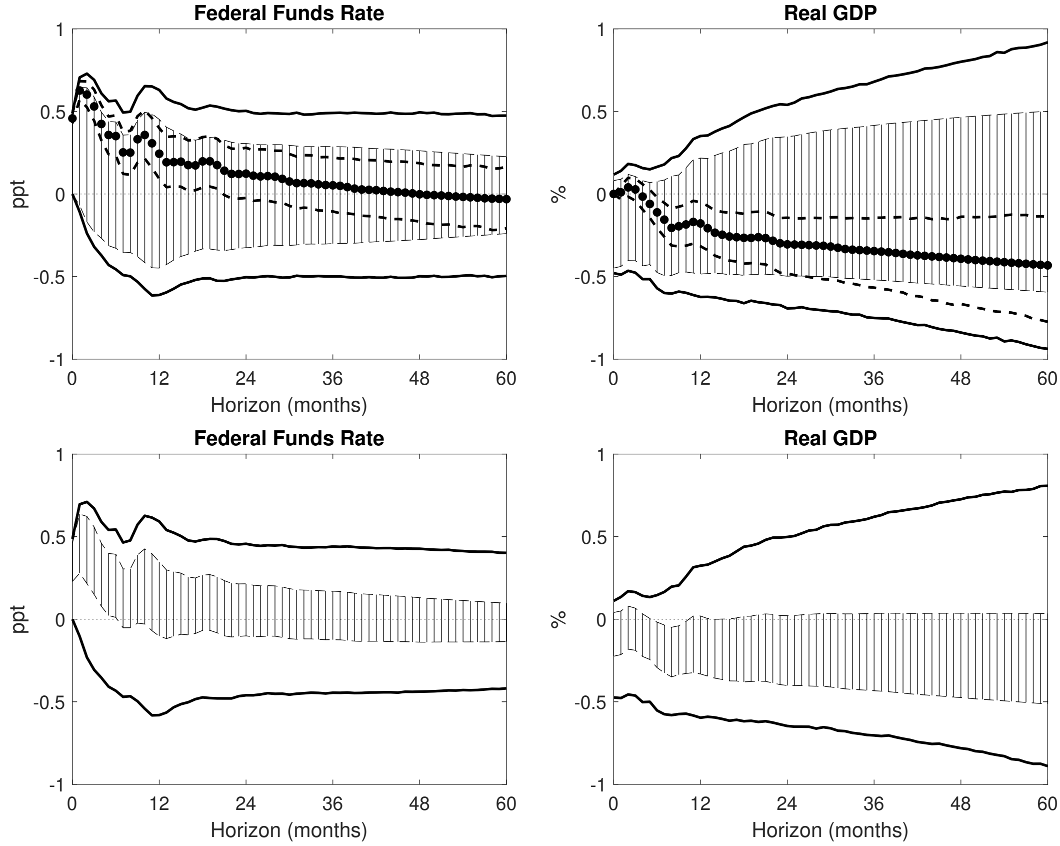
We assume there is uncertainty over the set of identifying restrictions. For the sake of illustration, we place equal weights on the two sets of restrictions. GKV discuss when the prior model probabilities are updated by the data. In particular, the prior model probabilities are not updated when the models are 'indistinguishable' (i.e., they admit an identical reduced-form defined on a common parameter space) and share a common prior for $\boldsymbol{\phi}$, where the prior for $\boldsymbol{\phi}$ is the notional reduced-form prior truncated to the region with nonempty identified set. Under the set-identifying restrictions considered, the identified set is never empty at any value of $\boldsymbol{\phi}$ supported by the reduced-form prior. The identified set under the point-identifying restrictions (which is a singleton) is also never empty. The two models share a common reduced-form prior and the posterior model probabilities are therefore equal to the prior model probabilities. The set of posteriors allowing for uncertainty over the identifying restrictions is then given by the simple average of the set of posteriors under the set-identifying restrictions and the single posterior under the point-identifying restrictions.

The top panels of Figure 4.2 plot the posterior means and 95 per cent highest posterior density intervals under the point-identifying restrictions alongside the robust Bayesian output under the set-identifying restrictions (i.e., the robust Bayesian output plotted in Figure 4.1). Under the point-identifying restrictions, output falls with high posterior probability. The bottom panels plot the model-averaged set of posterior means and robust credible intervals. The set of posterior means under the set-identifying restrictions has been shrunk towards the posterior mean under the point-identifying restrictions; since the posterior model probabilities are equal, the lower (upper) bound of the set of posterior means is the average of the lower (upper) bound of the set of posterior means in the set-identified model and the posterior mean in the point-identified model. As a result, the set of posterior means for the output response now excludes zero at horizons of less than one year. Nevertheless, the robust credible intervals contain zero at all horizons and the posterior lower probability that the output response is negative at the one-year horizon is only 50 per cent. Equally weighting models identified using the set-identifying restrictions in ACR and classic recursive restrictions thus provides fairly weak evidence that output falls following a monetary policy shock once one allows for ambiguity over the unrevisable prior in the set-identified model.

This approach can also be used to back out the prior model probabilities that would lead to particular posterior inferences. For example, one would need to place a prior probability of at least 0.55 on the point-identified model for the model-averaged set of posterior means to unambiguously imply a negative output response at the one-year horizon.

KL-neighborhood (GKU). To illustrate the approach in GKU, we treat the normal-inverse-Wishart prior for $\boldsymbol{\phi}$ and the conditionally uniform prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ as a benchmark prior and conduct a posterior sensitivity exercise. The conditionally uniform prior for $\mathbf{Q}$ given $\boldsymbol{\phi}$ implies a benchmark (conditional) prior for the impulse response of interest, η. We consider perturbations of the prior for η within a neighborhood of the benchmark prior. The size of the neighborhood is determined by the KL distance, λ. We consider different values of λ and document how posterior inference changes depending on the size of the neighborhood.

Figure 4.2: Impulse Responses to a Monetary Policy Shock – Uncertain Identification

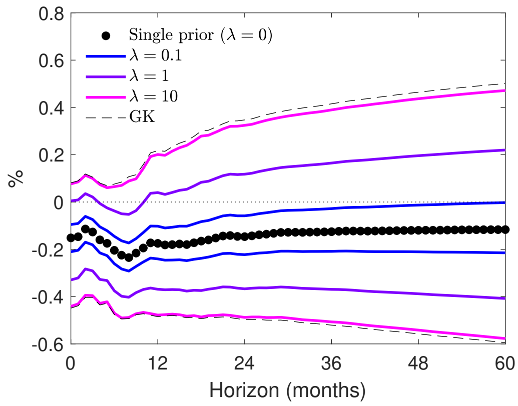


Notes: Top panels plot robust Bayesian output under set-identifying restrictions and standard Bayesian output under point-identifying restrictions. Bottom panels plot robust Bayesian output from GKV, assuming equal prior probabilities on the two sets of restrictions. Circles and dashed lines are, respectively, posterior means and 95 per cent (pointwise) highest posterior density intervals under point-identifying restrictions. Vertical bars are sets of posterior means and solid lines are 95 per cent (pointwise) robust credible regions.

At each draw of $\boldsymbol{\phi}$ from its posterior, $\{\boldsymbol{\phi}_m\}_{m=1}^{M}$, we obtain N draws of $\mathbf{Q}$ from the conditionally uniform distribution over the identified set. We transform these draws into impulse-response space, so we have a set of draws of η from the benchmark prior at each draw of $\boldsymbol{\phi}$, $\{\eta_{mi}\}_{i=1}^{N}$. We then approximate the Lagrange multipliers, $\kappa_\lambda^l(\boldsymbol{\phi})$ and $\kappa_\lambda^u(\boldsymbol{\phi})$, using an interior-point algorithm implemented within Matlab's 'fmincon' optimizer. After computing the Lagrange multipliers, we compute the set of posterior means using the sample analogue of (4.56) (i.e., with integration replaced by averaging over the draws of η and $\boldsymbol{\phi}$). For the purposes of illustration, we conduct the sensitivity exercise under different values of $\lambda \in \{0.1, 1, 10\}$.

Figure 4.3 presents the set of posterior means of the output response to a mon-

Figure 4.3: Impulse Responses to a Monetary Policy Shock – Posterior Sensitivity Analysis



Notes: Circles are posterior means under the benchmark conditional prior, colored solid lines are bounds of sets of posterior means for different values of the Kullback-Leibler distance λ and dashed lines are bounds of sets of posterior means using the approach from GK (i.e., from the right panel of Figure 4.1).

etary policy shock at each horizon of interest and for each value of λ considered. For comparison, the figure also plots the posterior mean under the benchmark prior (which is equivalent to the case where $\lambda = 0$) and the set of posterior means obtained using the approach from GK described above (which is equivalent to the case where $\lambda \to \infty$). When $\lambda = 0.1$, the set of priors is constrained to lie within a relatively small KL distance of the benchmark prior, and the resulting set of posterior means lies entirely below zero at all horizons considered; in other words, for priors within a relatively small neighborhood of the benchmark prior, the set of posterior means supports the conclusion that output falls following a positive monetary policy shock. As λ increases, we allow for priors further from the benchmark prior and the set of posterior means expands. At $\lambda = 1$, the set of posterior means excludes zero at only some horizons shorter than one year, and at $\lambda = 10$ the set of posterior means includes zero at all horizons. For large values of λ, the set of priors grows to include all priors consistent with the identifying restrictions. Consequently, as λ increases, the set of posterior means converges towards that obtained using the approach from GK, which allows for full ambiguity over the set of priors consistent with the identifying restrictions.

4.8 Conclusion

We overviewed how robust Bayesian analysis provides useful tools for Bayesian econometricians and decision-makers who want to assess the sensitivity of inferences and statistical decisions to the choice of a prior. The main idea is to construct a set of priors, which in turn delivers: 1) a set of posterior quantities that can be used for inference; and/or 2) an optimal statistical decision for ambiguity-averse decision-makers. We discussed how the sensitivity concerns are particularly salient in set-identified structural models due to the fact that a component of the prior is not revised by the data even in large samples. We reviewed different ways to construct the set of priors and discussed in detail how to implement the methods in macroeconometric applications using set-identified SVARs.

Chapter 5

Robust Bayesian Inference in Proxy SVARs

5.1 Introduction

Proxy structural vector autoregressions (SVARs) are an increasingly popular method for estimating the dynamic causal effects of macroeconomic shocks.[1] The key identifying assumption in the proxy SVAR is that there exists one or more variables external to the SVAR – 'proxies' or 'external instruments' – that are correlated with particular structural shocks (i.e., 'relevant') and uncorrelated with all other structural shocks (i.e., 'exogenous'). The impulse responses to a single structural shock can be point-identified when a single proxy is correlated with that structural shock and uncorrelated with all other structural shocks (Stock (2008)). Mertens and Ravn (2013) (henceforth MR13) develop a proxy SVAR with multiple proxies for multiple structural shocks and show that point identification of the impulse responses to these shocks requires zero restrictions on the structural parameters in addition to the zero restrictions implied by exogeneity of the proxies. Other papers that use multiple proxies to identify multiple structural shocks given additional point-identifying restrictions include Lunsford (2015) and Mertens and Montiel Olea (2018). The additional restrictions required to achieve point iden-

[1] See, for example, Stock and Watson (2012, 2016, 2018), Mertens and Ravn (2013, 2014, 2019), Gertler and Karadi (2015), Lunsford (2015), Ramey (2016), Caldara and Kamps (2017), Mertens and Montiel Olea (2018), Angelini and Fanelli (2019), Caldara and Herbst (2019), Jentsch and Lunsford (2019), Bahaj (2020), Drautzburg (2020), Arias et al. (2021) and Montiel Olea *et al.* (2021).

tification may not always have a theoretically sound motivation. Consequently, there may be interest in assessing the robustness of the analysis to relaxing these additional restrictions, which would result in set identification.

The majority of the literature that makes use of proxy SVARs conducts inference in the frequentist setting. A notable exception is Arias et al. (2021) (henceforth ARW21), who develop algorithms for Bayesian inference that are applicable under set identification. Bayesian inference may be appealing because it allows the researcher to use prior information about the model's parameters and, under set identification, it may be computationally more convenient than a frequentist approach. This is perhaps why, since Uhlig (2005), the dominant inferential approach in set-identified SVARs has been Bayesian.[2] However, under set identification, posterior inference is sensitive to the choice of prior over the set-identified parameters, even asympotically (Poirier (1998)), and Bayesian credible intervals do not asymptotically coincide with frequentist confidence intervals (Moon and Schorfheide (2012)). Moreover, in the context of SVARs, Baumeister and Hamilton (2015) show that even priors that are 'uniform' over a set-identified parameter may be informative about the objects of interest, such as impulse responses.

To address these issues, Giacomini and Kitagawa (2021) (GK) propose an approach to Bayesian inference in set-identified models that is robust to the choice of prior over the set-identified parameters. The approach considers the class of all priors over the model's set-identified parameters that are consistent with the identifying restrictions. This generates a class of posteriors, which can be summarised by reporting the set of posterior means (an estimator of the identified set) and a robust credible region. GK provide conditions under which these quantities have valid frequentist interpretations and they apply their approach to SVARs in which the impulse responses are set-identified by imposing sign and zero restrictions.

In this paper we extend the approach of GK to set-identified proxy SVARs. Following MR13 and ARW21, we consider the case where there are $k < n$ proxies

[2]Gafarov *et al.* (2018) and Granziera *et al.* (2018) develop frequentist inferential tools in set-identified SVARs. We are unaware of papers that conduct frequentist inference in set-identified proxy SVARs.

that are correlated with k structural shocks (a 'relevance' condition) and are uncorrelated with the remaining $n-k$ shocks (an 'exogeneity' condition), where n is the dimension of the SVAR. If $n > 3$ and $1 < k < n-1$, the impulse responses to all structural shocks are set-identified in the absence of further zero restrictions on the structural parameters. For other values of n and k, it may be the case that impulse responses to particular structural shocks are point-identified, while other impulse responses are set-identified. We focus on cases where the impulse responses of interest are set-identified.

This paper makes several new contributions relative to GK. First, we provide conditions under which our procedure is guaranteed to have a valid frequentist interpretation in proxy SVARs. These results do not follow directly from those in GK, and are tailored to the different structure of the problem in proxy SVARs. Second, we show that, in the presence of weak proxies, both the frequentist and the Bayesian approach no longer provide asymptotically valid inference about the identified set: the estimators of the bounds of the identified set are not consistent, they converge to non-degenerate and data-dependent distributions, and these distributions are different for the frequentist and the Bayesian approach (implying a failure of the Bernstein-von Mises property that we prove holds under strong proxies). Third, we show how to conduct posterior inference not only about the impulse responses, but also about the forecast error variance decomposition (FEVD), which is the relative contribution of a particular structural shock to the unexpected variation in a particular variable over some horizon.[3] Finally, we provide an algorithm for computing impulse responses to a unit shock (as opposed to a standard-deviation shock), which are often considered in the proxy-SVAR literature.

As in ARW21, our algorithms allow for zero and sign restrictions on the covariances between the proxies and the structural shocks in addition to the zero restrictions implied by the exogeneity assumption. These types of restrictions are likely to be justifiable in applications, given that the proxies are typically constructed with

[3] Plagborg-Møller and Wolf (in press) develop frequentist procedures for conducting inference about the FEVD in a general semiparametric moving average model when there are valid external instruments available. The setting that they consider allows for cases where the FEVD is set-identified.

the purpose of measuring a particular structural shock. An example of a zero restriction would be to assume that, among the k structural shocks that are assumed to be correlated with the k proxies, a particular structural shock is uncorrelated with a particular proxy. Examples of sign restrictions are when a particular proxy is positively correlated with a particular structural shock, or when the covariance between a particular proxy and a particular structural shock is larger than the covariance between that proxy and another structural shock.[4] Additionally, our algorithms allow for restrictions of the kind considered in GK, including 'short-run' zero restrictions (as in Sims (1980) and Christiano *et al.* (1999)), 'long-run' zero restrictions (as in Blanchard and Quah (1989)), sign restrictions on impulse responses (as in Uhlig (2005)), and zero or sign restrictions on the matrix determining the contemporaneous relationships among the endogenous variables (as in Arias et al. (2019)). By extending and adapting the algorithms in GK to allow for identification using proxy variables alongside standard zero and sign restrictions, we provide a general and flexible tool for empirical researchers to relax potentially controversial point-identifying restrictions without having to adopt an unrevisable prior.

Some existing approaches to Bayesian inference in proxy SVARs place priors directly on the model's structural parameters. For example, ARW21 place a normal-generalised-normal conjugate prior over the proxy SVAR's structural parameters and propose algorithms for drawing from the resulting normal-generalised-normal posterior. More generally, Baumeister and Hamilton (2015, 2018, 2019) advocate placing priors on the structural parameters of an SVAR, because these parameters can have economic interpretations that facilitate prior elicitation. A problem with this approach in set-identified models is that the prior implicitly incorporates a component that is unrevisable by the data. Our approach overcomes this problem by decomposing the prior over the structural parameters into a revisable prior over reduced-form parameters and an unrevisable prior over the orthonormal matrix that

[4]The first type of sign restriction is considered by Ludvigson *et al.* (2018) and Piffer and Podstawski (2018) in the frequentist setting and by Braun and Brüggemann (2017) and ARW21 in the Bayesian setting, while the second type is considered by Braun and Brüggemann (2017), Piffer and Podstawski (2018) and ARW21. Braun and Brüggemann (2017) and Piffer and Podstawski (2018) assume that the proxy is correlated with all structural shocks (i.e., there are no exogeneity restrictions); it would be straightforward to implement this setup under our approach.

maps VAR innovations into structural shocks (see, for example, Uhlig (2005)). We then allow for multiple priors for this matrix, which delivers inference that is robust to the choice of unrevisable prior. We see our approach as being complementary to existing Bayesian approaches. In particular, we suggest reporting output based on the multiple-prior robust Bayesian approach together with output from the single-prior Bayesian posterior to document the sensitivity of posterior inference to the choice of unrevisable prior.[5]

It is well-known that frequentist inference in the linear instrumental-variables model is non-standard when the instruments are weakly correlated with the included endogenous variables (e.g., Stock *et al.* (2002)). Similar problems arise in the proxy SVAR when the proxies are weakly correlated with the structural shocks. In the case where there is one proxy for one structural shock, Lunsford (2015) shows that the estimator of the impulse response is inconsistent when the proxy is weak, and he derives a test for the presence of a weak proxy. Montiel Olea *et al.* (2021) show that standard asymptotic (delta-method) inference about the objects of interest in the proxy SVAR is invalid when the proxy is weak, and they derive a weak-instrument-robust confidence interval for the impulse response. As noted in Caldara and Herbst (2019), from the standpoint of Bayesian inference, having a weak proxy does not invalidate posterior inference in the sense that one still obtains (numerical approximations of) the exact finite-sample posterior distributions of the objects of interest. However, practitioners may be interested in the asymptotic frequentist properties of Bayesian inferential procedures. For example, Bayesians may be better able to credibly communicate their results to frequentist audiences when the Bayesian inferential procedure is asymptotically equivalent to a frequentist procedure. Accordingly, we investigate the asymptotic properties of our robust Bayesian procedure in the presence of weak instruments. Using a simple analytical example, we show that our robust Bayesian procedure does not provide valid frequentist inference about the identified set under weak-proxy asymptotics, which contrasts with the results in Kline and Tamer (2016) and GK. To the best of our knowledge, this is the first

[5] An alternative approach is to consider variation in the prior within some neighbourhood around a benchmark prior, as in Giacomini *et al.* (2019).

paper to provide formal results on the interplay between set identification and weak identification.

We illustrate our procedure by considering the analysis in MR13, which is also discussed in Jentsch and Lunsford (2019) and Mertens and Ravn (2019). MR13 use series of plausibly exogenous, unanticipated changes in average personal and corporate income tax rates in the United States as proxies for structural shocks to these tax rates to identify the effects of fiscal shocks on macroeconomic variables. Since there are two proxies for two structural shocks, the impulse responses to these shocks are set-identified in the absence of additional zero restrictions. MR13 impose a zero restriction in addition to those implied by exogeneity of the proxies, which yields point identification. The additional restriction is a causal ordering, which restricts the direct contemporaneous response of one tax rate to the other. This assumption could be violated if, for instance, there are constraints that impinge on the ability of the government to change tax rates independently of one another. MR13 assess the robustness of the results to imposing the additional restriction by considering two alternative causal orderings of the tax rates within the proxy SVAR. Our approach extends and formalizes this robustness analysis by providing an estimator of the set of impulse responses compatible with relaxing the additional zero restriction and replacing it with a set of – arguably weaker – sign restrictions. We compare the results under our multiple-prior Bayesian approach to those obtained under a single prior to assess the role of prior choice in driving posterior inference.

Outline. The remainder of the paper is structured as follows. Section 5.2 describes our robust Bayesian inferential framework for set-identified proxy SVARs. Section 5.3 provides results on the frequentist properties of this approach and explores how weak proxies affect posterior inference asymptotically. Section 5.4 details the numerical algorithms used to implement the approach. Section 5.5 contains the empirical application and Section 5.6 concludes.

Generic notation: For the matrix $\mathbf{X}$, $\text{vec}(\mathbf{X})$ is the vectorisation of $\mathbf{X}$ and $\text{vech}(\mathbf{X})$ is the half-vectorisation of $\mathbf{X}$ (when $\mathbf{X}$ is symmetric). $\mathbf{e}_{i,n}$ is the ith column of the $n \times n$ identity matrix, $\mathbf{I}_n$. $\mathbf{0}_{n\times m}$ is a $n \times m$ matrix of zeros. $\|.\|$ is the Euclidean norm.

$\mathscr{S}^{n-1}$ is the unit sphere in $\mathbb{R}^n$.

5.2 Framework

5.2.1 The SVAR

Let $\mathbf{y}_t$ be an $n \times 1$ vector of endogenous variables following the SVAR(p) process:

$$\mathbf{A}_0\mathbf{y}_t = \sum_{l=1}^{p} \mathbf{A}_l\mathbf{y}_{t-l} + \boldsymbol{\varepsilon}_t, \quad t = 1,...,T,$$

where $\mathbf{A}_0$ has positive diagonal elements (a sign normalisation) and is invertible, and $\boldsymbol{\varepsilon}_t$ are structural shocks with $\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{I}_n$. The initial conditions $(\mathbf{y}_{1-p},...,\mathbf{y}_0)$ are given. We omit exogenous regressors (such as a constant) for simplicity of exposition, but these are straightforward to include. Letting $\mathbf{x}_t = (\mathbf{y}_{t-1}',\ldots,\mathbf{y}_{t-p}')'$ and $\mathbf{A}_+ = (\mathbf{A}_1,\ldots,\mathbf{A}_p)$, we can rewrite the SVAR(p) as

$$\mathbf{A}_0\mathbf{y}_t = \mathbf{A}_+\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1,...,T. \tag{5.1}$$

$(\mathbf{A}_0,\mathbf{A}_+)$ are the structural parameters. The reduced-form VAR(p) representation is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u}_t, \quad t = 1,...,T,$$

where $\mathbf{B} = (\mathbf{B}_1,\ldots,\mathbf{B}_p)$, $\mathbf{B}_l = \mathbf{A}_0^{-1}\mathbf{A}_l$ for $l = 1,\ldots,p$, and $\mathbf{u}_t = \mathbf{A}_0^{-1}\boldsymbol{\varepsilon}_t$ with $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$. $(\mathbf{B},\boldsymbol{\Sigma})$ are the reduced-form parameters. We assume that $\mathbf{B}$ is such that the VAR(p) can be inverted into an infinite-order vector moving average (VMA(∞)) model.[6]

To facilitate computing the identified set of the objects of interest, we reparameterise the model into its 'orthogonal reduced form':

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \boldsymbol{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_t, \quad t = 1,...,T,$$

where $\boldsymbol{\Sigma}_{tr}$ is the lower-triangular Cholesky factor of $\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{\Sigma}_{tr}\boldsymbol{\Sigma}_{tr}' = \boldsymbol{\Sigma}$) with di-

[6]The VAR(p) is invertible into a VMA(∞) process when the eigenvalues of the companion matrix lie inside the unit circle. See Hamilton (1994) or Kilian and Lütkepohl (2017).

agonal elements normalised to be non-negative, $\mathbf{Q} \in \mathscr{O}(n)$ is an $n \times n$ orthonormal matrix and $\mathscr{O}(n)$ is the set of all such matrices. The parameterisations are related through the mapping $\mathbf{B} = \mathbf{A}_0^{-1}\mathbf{A}_+$, $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$ and $\mathbf{Q} = \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{A}_0^{-1}$, or $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$ and $\mathbf{A}_+ = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{B}$. The sign normalisation that the diagonal elements of $\mathbf{A}_0$ are nonnegative therefore corresponds to the restriction that $\text{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times 1}$.

The VMA(∞) representation of the model is

$$\mathbf{y}_t = \sum_{h=0}^{\infty} \mathbf{C}_h \mathbf{u}_{t-h} = \sum_{h=0}^{\infty} \mathbf{C}_h \boldsymbol{\Sigma}_{tr} \mathbf{Q} \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T,$$

where $\mathbf{C}_h$ is the hth term in $(\mathbf{I}_n - \sum_{l=1}^{p} \mathbf{B}_l L^l)^{-1}$ and L is the lag operator. The (i, j)th element of the matrix $\mathbf{C}_h\boldsymbol{\Sigma}_{tr}\mathbf{Q}$, which we denote by $\eta_{i,j,h}$, is the impulse response of the ith variable to the jth structural shock at the hth horizon:

$$\eta_{i,j,h} = \mathbf{e}_{i,n}' \mathbf{C}_h \boldsymbol{\Sigma}_{tr} \mathbf{Q} \mathbf{e}_{j,n} = \mathbf{c}_{i,h}' \mathbf{q}_j, \tag{5.2}$$

where $\mathbf{c}_{i,h}' \equiv \mathbf{e}_{i,n}'\mathbf{C}_h\boldsymbol{\Sigma}_{tr}$ is the ith row of $\mathbf{C}_h\boldsymbol{\Sigma}_{tr}$ and $\mathbf{q}_j \equiv \mathbf{Q}\mathbf{e}_{j,n}$ is the jth column of $\mathbf{Q}$.

Another object that is also often of interest in analyses using (proxy) SVARs is the FEVD. Under quadratic loss, the optimal h-step-ahead forecast of $\mathbf{y}_t$ given information available at time t is $\mathbb{E}\left(\mathbf{y}_{t+h}|\mathscr{F}_t\right) = \sum_{k=0}^{\infty} \mathbf{C}_{h-k}\mathbf{u}_{t-k}$. The h-step-ahead forecast error is then $\mathbf{y}_{t+h} - \mathbb{E}\left(\mathbf{y}_{t+h}|\mathscr{F}_t\right) = \sum_{k=0}^{h-1} \mathbf{C}_k \mathbf{u}_{t+h-k} = \sum_{k=0}^{h-1} \mathbf{C}_k\boldsymbol{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_{t+h-k}$. It follows that the forecast error variance of $y_{i,t+h}$ is $\text{var}(y_{i,t+h}|\mathscr{F}_t) = \sum_{k=0}^{h-1} \mathbf{c}_{i,k}'\mathbf{c}_{i,k}$. The contribution of the jth structural shock to the forecast error variance of the ith variable at the hth horizon is $\text{var}(y_{i,t+h}|\mathscr{F}_t, \boldsymbol{\varepsilon}_{-j,t+1}, \ldots, \boldsymbol{\varepsilon}_{-j,t+h}) = \sum_{k=0}^{h-1} \mathbf{c}_{i,k}'\mathbf{q}_j\mathbf{q}_j'\mathbf{c}_{i,k}$, where $\boldsymbol{\varepsilon}_{-j,t} = \{\varepsilon_{i,t} : i \neq j \wedge i = 1, \ldots, n\}$. The contribution of the jth structural shock to the forecast error variance of the ith variable at the hth horizon as a fraction of the total forecast error variance is then

$$\text{FEVD}_{i,j,h} = \frac{\sum_{k=0}^{h-1} \mathbf{c}_{i,k}'\mathbf{q}_j\mathbf{q}_j'\mathbf{c}_{i,k}}{\sum_{k=0}^{h-1} \mathbf{c}_{i,k}'\mathbf{c}_{i,k}}. \tag{5.3}$$

5.2.2 Identification using proxies

In the absence of identifying restrictions, the structural parameters – and any function of these parameters, such as the impulse responses or FEVD – are set-identified. Since any $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$ satisfies $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$, the identified set for $\mathbf{A}_0$ is $\{\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1} : \mathbf{Q} \in \mathscr{O}(n)\}$. Imposing identifying restrictions restricts $\mathbf{Q}$ to lie in a subspace $\mathscr{Q}$ of $\mathscr{O}(n)$, which shrinks the identified set.

The key identifying assumption in the proxy SVAR is that there are variables external to the SVAR that are correlated with particular structural shocks and uncorrelated with all other structural shocks. Let $\boldsymbol{\varepsilon}_{(i:j),t} = (\varepsilon_{i,t}, \varepsilon_{i+1,t}, ..., \varepsilon_{j-1,t}, \varepsilon_{j,t})'$ for $i < j$. Assume that $\mathbf{m}_t$ is a $k \times 1$ vector of proxies (with $k < n$) that are correlated with the last k structural shocks, so $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(n-k+1:n),t}) = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a full-rank $k \times k$ matrix. Further, assume that $\mathbf{m}_t$ is uncorrelated with the first $n-k$ structural shocks, so $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(1:n-k),t}) = \mathbf{0}_{k\times(n-k)}$. The first condition is commonly referred to as the 'relevance' condition and the second as the 'exogeneity' condition. We assume that $\mathbf{m}_t$ is generated by the process

$$\boldsymbol{\Gamma}_0 \mathbf{m}_t = \boldsymbol{\Lambda}\boldsymbol{\varepsilon}_t + \sum_{l=1}^{p_m} \boldsymbol{\Gamma}_l \mathbf{m}_{t-l} + \boldsymbol{\nu}_t, \quad t = 1, ..., T, \tag{5.4}$$

where: $\boldsymbol{\Gamma}_l$, $l = 0, ..., p_m$, is a $k \times k$ matrix with $\boldsymbol{\Gamma}_0$ invertible; $\boldsymbol{\Lambda}$ is a $k \times n$ matrix; and the initial conditions $(\mathbf{m}_{1-p_m}, \ldots, \mathbf{m}_0)$ are given. We assume that $(\boldsymbol{\varepsilon}_t', \boldsymbol{\nu}_t')'|\mathscr{F}_{t-1} \sim N(\mathbf{0}_{(n+k)\times 1}, \mathbf{I}_{n+k})$, where $\mathscr{F}_{t-1}$ is the information set at time $t-1$, which includes the lags of $\mathbf{y}_t$ and $\mathbf{m}_t$. The assumption about the joint distribution of $(\boldsymbol{\varepsilon}_t, \boldsymbol{\nu}_t)$ implies that $\boldsymbol{\nu}_t|\mathscr{F}_{t-1}, \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{k\times 1}, \mathbf{I}_k)$. This process is an SVAR(p_m) in $\mathbf{m}_t$ where the structural shocks $\boldsymbol{\varepsilon}_t$ are included as exogenous variables. The process implies that the proxies contain information about the structural shocks after allowing for possible serial correlation in the proxies.[7] The information content of each proxy for each structural shock is jointly determined by the matrices $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Lambda}$. This setup allows for the number of lags of $\mathbf{m}_t$ in the SVAR for $\mathbf{m}_t$ to differ to the number of lags of

[7] We could also allow for up to p lags of $\mathbf{y}_t$ to appear in (5.4) without altering the reduced form for $\mathbf{m}_t$ or the restrictions on $\mathbf{Q}$ implied by proxy exogeneity (derived below).

$\mathbf{y}_t$ in the SVAR for $\mathbf{y}_t$.[8]

Given the distributional assumption on $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\nu}_t$, and the exogeneity and relevance assumptions, it follows from (5.4) that

$$\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda} = \left[\mathbf{0}_{k\times(n-k)}, \boldsymbol{\Psi}\right]. \tag{5.5}$$

Left-multiplying (5.4) by $\boldsymbol{\Gamma}_0^{-1}$ and substituting out $\boldsymbol{\varepsilon}_t$ using (5.1) yields

$$\mathbf{m}_t = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda}\mathbf{A}_0\mathbf{y}_t - \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda}\mathbf{A}_+\mathbf{x}_t + \sum_{l=1}^{p_m}\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_l\mathbf{m}_{t-l} + \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\nu}_t.$$

The reduced-form process for the proxies, which we refer to as the 'first-stage regression', is

$$\mathbf{m}_t = \mathbf{D}\mathbf{y}_t + \mathbf{G}\mathbf{x}_t + \sum_{l=1}^{p_m}\mathbf{H}_l\mathbf{m}_{t-l} + \mathbf{v}_t, \tag{5.6}$$

where: $\mathbf{D} = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda}\mathbf{A}_0$; $\mathbf{G} = -\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda}\mathbf{A}_+$; $\mathbf{H}_l = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_l$ for $l = 1,...,p_m$; and $\mathbf{v}_t = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\nu}_t$ with $\mathbb{E}(\mathbf{v}_t\mathbf{v}_t') = \boldsymbol{\Upsilon} = \boldsymbol{\Gamma}_0^{-1}(\boldsymbol{\Gamma}_0^{-1})'$. This is a VAR($p_m$) in $\mathbf{m}_t$ with exogenous variables $\mathbf{y}_t$ and $\mathbf{x}_t$. The first-stage regression should also include any exogenous variables (e.g., a constant) that are included in the SVAR for $\mathbf{y}_t$. Since $\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Lambda} = \mathbf{D}\mathbf{A}_0^{-1} = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q}$, we can write (5.5) as

$$\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}_t') = \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q} = \left[\mathbf{0}_{k\times(n-k)}, \boldsymbol{\Psi}\right]. \tag{5.7}$$

The (i,j)th element of this matrix is $\mathbf{e}_{i,k}'\mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_{j,n} = \mathbf{d}_i'\mathbf{q}_j$, where $\mathbf{d}_i' \equiv \mathbf{e}_{i,k}'\mathbf{D}\boldsymbol{\Sigma}_{tr}$ is the ith row of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$. The exogeneity assumption therefore generates linear restrictions on the first $n-k$ columns of $\mathbf{Q}$ given the reduced-form parameters $\mathbf{D}$ and $\boldsymbol{\Sigma}_{tr}$. The proxies satisfy the relevance assumption rank($\boldsymbol{\Psi}$) $= k$ if and only if rank($\mathbf{D}$) $= k$.

Let f_i be the number of equality restrictions on the ith column of $\mathbf{Q}$. Rubio-Ramírez *et al.* (2010) show that a necessary and sufficient condition for point identification of the structural parameters in an SVAR is that $f_i = n-i$ for $i = 1,...,n$.

[8]ARW21 specify a joint SVAR for $(\mathbf{y}_t', \mathbf{m}_t')'$ where zero restrictions rule out feedback from $\mathbf{m}_t$ to $\mathbf{y}_t$. This process also implies that the proxies contain information about the structural shocks and, under the exogeneity assumption, yields the same set of identifying zero restrictions that we derive below.

We focus on cases where $f_i \leq n-i$ for all $i = 1,...,n$, with strict inequality for at least one i, and where interest is in a particular set-identified object. Equations (5.2) and (5.3) imply that the impulse response and FEVD corresponding to the jth structural shock are point-identified if and only if the jth column of $\mathbf{Q}$ is point-identified. Assume for now that the only zero restrictions are those corresponding to the exogeneity assumption and that $n \geq 3$. Assume also that $\text{rank}(\mathbf{D}) = k$, so the relevance condition holds. If $k = 1$, then $f_i = 1$ for $i = 1,...,n-1$ and $f_n = 0$. In this case, the first $n-1$ columns of $\mathbf{Q}$ are set-identified and $\mathbf{q}_n$ is point-identified.[9] If $k = n-1$, then $f_1 = n-1$ and $f_i = 0$ for $i = 2,...,n$. In this case, $\mathbf{q}_1$ is point-identified and $\mathbf{q}_i$, $i = 2,...,n$, is set-identified.[10] For $1 < k < n-1$, all columns of $\mathbf{Q}$ are set-identified.[11]

As in ARW21, we allow for additional equality and sign restrictions on elements of $\boldsymbol{\Psi}$. An example of an equality restriction is that the first proxy variable (m_{1t}) is not only uncorrelated with the first $n-k$ structural shocks, but is also uncorrelated with one of the last k structural shocks (e.g., $\mathbb{E}(m_{1t}\varepsilon_{(n-k+1),t}) = 0$). This type of restriction is a linear equality restriction on a single column of $\mathbf{Q}$. An example of a sign restriction is that the covariance between the first proxy and one of the last k structural shocks is nonnegative (e.g., $\mathbb{E}(m_{1t}\varepsilon_{nt}) \geq 0$), which is a linear inequality restriction on a single column of $\mathbf{Q}$. Another example is that the covariance between a particular proxy and a particular structural shock is greater than or equal to the covariance between that proxy and another structural shock, which is a linear inequality restriction on two columns of $\mathbf{Q}$; for example, $\mathbb{E}(m_{1t}\varepsilon_{nt}) \geq \mathbb{E}(m_{1t}\varepsilon_{n-1,t})$ implies that $\mathbf{d}_1'(\mathbf{q}_n - \mathbf{q}_{n-1}) \geq 0$.

Our approach also allows for other restrictions commonly used in SVARs, such as zero restrictions on $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$, $\mathbf{A}_0^{-1} = \boldsymbol{\Sigma}_{tr}\mathbf{Q}$ or the long-run cumulative im-

[9] The exogeneity restrictions imply that $\mathbf{d}_1'\mathbf{q}_i = 0$ for $i = 1,\ldots,n-1$. Since the columns of an orthonormal matrix are orthogonal and have unit length, $\mathbf{q}_n = \pm\mathbf{d}_1/\|\mathbf{d}_1\|$. The sign normalisation pins down the sign of $\mathbf{q}_n$.

[10] The exogeneity restrictions imply that $\mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{q}_1 = 0$, where $\mathbf{D}$ is a $(n-1)\times n$ matrix. Under the relevance assumption, $\text{rank}(\mathbf{D}\boldsymbol{\Sigma}_{tr}) = n-1$ and the nullspace of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$ is of dimension one by the rank-nullity theorem. $\mathbf{q}_1$ is therefore a unit-length vector in the (one-dimensional) nullspace of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$, which is uniquely determined given the sign normalisation.

[11] The result for $k = 1$ corresponds to Corollary 2 in ARW21. The results for $k = n-1$ and $1 < k < n-1$ follow from their Proposition 2.

pulse response $CIR^{\infty} = (\mathbf{I}_n - \sum_{l=1}^{p} \mathbf{B}_l)^{-1} \blacksquare_{tr} \mathbf{Q}$, and sign restrictions on the impulse responses or $\mathbf{A}_0$.

5.2.3 Robust Bayesian inference

We assume for now that the object of interest is the impulse response $\eta_{i,j,h}$, although the discussion in this section also applies to the FEVD or any other scalar-valued function of the structural parameters. Given the formulation of the exogeneity restrictions and any additional zero or sign restrictions as restrictions on the columns of $\mathbf{Q}$, robust Bayesian inference about the identified set for $\eta_{i,j,h}$ proceeds as in GK. We summarise the salient features of this approach here.

Collect the coefficients on $\mathbf{x}_t$ and $\mathbf{m}_t$ in (5.6) as

$$\mathbf{J} = [\text{vec}(\mathbf{G})', \text{vec}(\mathbf{H}_1)', \ldots, \text{vec}(\mathbf{H}_{p_m})']'.$$

We denote the proxy-SVAR reduced-form parameters as

$$\boldsymbol{\phi} = (\text{vec}(\mathbf{B})', \text{vech}(\boldsymbol{\Sigma})', \text{vec}(\mathbf{D})', \mathbf{J}', \text{vech}(\boldsymbol{\Upsilon})')' \in \boldsymbol{\Phi}.$$

Since the zero restrictions are linear equality restrictions on single columns of $\mathbf{Q}$ and are otherwise functions only of the reduced-form parameters, we can represent them in the general form

$$\mathbf{F}(\boldsymbol{\phi}, \mathbf{Q}) = \begin{bmatrix} \mathbf{F}_1(\boldsymbol{\phi})\mathbf{q}_1 \\ \vdots \\ \mathbf{F}_n(\boldsymbol{\phi})\mathbf{q}_n \end{bmatrix} = \mathbf{0}_{(\sum_{i=1}^{n} f_i) \times 1},$$

where $\mathbf{F}_i(\boldsymbol{\phi})$ is an $f_i \times n$ matrix that stacks the coefficient vectors of the zero restrictions constraining $\mathbf{q}_i$. If the zero restrictions do not constrain $\mathbf{q}_i$, $\mathbf{F}_i(\boldsymbol{\phi})$ does not exist and $f_i = 0$. We represent the sign restrictions as $S(\boldsymbol{\phi}, \mathbf{Q}) \geq \mathbf{0}_{s \times 1}$, where s is the number of sign restrictions (excluding the sign normalisation). If there are no sign restrictions, then $S(\boldsymbol{\phi}, \mathbf{Q})$ does not exist and $s = 0$.

To simplify implementation of the robust Bayesian inferential approach, we order the variables in $\mathbf{y}_t$ to satisfy Definition 1.

Definition 1 (Ordering of Variables): Given an ordering of the proxies in $\mathbf{m}_t$, order the variables in $\mathbf{y}_t$ so that f_i satisfies $f_1 \geq f_2 \geq \ldots \geq f_n \geq 0$. In case of ties, if the impulse response of interest is to the j^*th structural shock, order the j^*th variable first. That is, set $j^* = 1$ when no other column of $\mathbf{Q}$ has a larger number of restrictions than $\mathbf{q}_{j^*}$. If $j^* \geq 2$, order the variables so that $f_{j^*-1} > f_{j^*}$.

This ordering convention is used when iteratively constructing columns of $\mathbf{Q}$ satisfying the zero restrictions. It is an extension of the ordering convention used by Rubio-Ramírez *et al.* (2010) in the point-identified setting to allow for set identification due to sign restrictions and underidentifying zero restrictions, and mirrors Definition 3 in GK. The ordering convention uniquely determines j^*, but the ordering of the remaining variables will not be unique when $f_i = f_k$ for some $i, k \neq j^*$. However, re-ordering the remaining variables will have no effect on the results of our algorithms, as long as the ordering convention is satisfied. The following example illustrates how to order the variables to satisfy Definition 1 and, given the ordering, how the matrices of restrictions are constructed.

Example 5.2.1. Consider a proxy SVAR for (c_t, i_t, y_t, π_t), where c_t is consumption growth, i_t is investment growth, y_t is output growth and π_t is inflation. Assume that there exist two proxy variables, $\mathbf{m}_t = (m_{c,t}, m_{i,t})'$, which are correlated with the structural shocks $\varepsilon_{c,t}$ and $\varepsilon_{i,t}$, and are uncorrelated with $\varepsilon_{y,t}$ and $\varepsilon_{\pi,t}$. In the absence of additional zero restrictions, all impulse responses are set-identified. If the impulse response of interest is to $\varepsilon_{i,t}$, an ordering of the variables that satisfies Definition 1 is (y_t, π_t, i_t, c_t), with $(f_1, f_2, f_3, f_4) = (2,2,0,0)$ and $j^* = 3$. If, instead, the impulse response of interest is to $\varepsilon_{\pi,t}$, an ordering of the variables that satisfies Definition 1 is (π_t, y_t, i_t, c_t), with $(f_1, f_2, f_3, f_4) = (2,2,0,0)$ and $j^* = 1$. In both cases, $\mathbf{F}_1(\boldsymbol{\phi}) = \mathbf{F}_2(\boldsymbol{\phi}) = \mathbf{D}\boldsymbol{\Sigma}_{tr}$ is a 2×4 matrix.

In the case where $j^* = 3$, consider the additional sign restrictions $\mathbb{E}(m_{c,t}\varepsilon_{c,t}) \geq 0$, $\mathbb{E}(m_{i,t}\varepsilon_{i,t}) \geq 0$, $\mathbb{E}(m_{c,t}\varepsilon_{c,t}) \geq \mathbb{E}(m_{c,t}\varepsilon_{i,t})$ and $\mathbb{E}(m_{i,t}\varepsilon_{i,t}) \geq \mathbb{E}(m_{i,t}\varepsilon_{c,t})$. The matrix of sign restrictions can be represented as

$$\mathbf{S}(\boldsymbol{\phi},\mathbf{Q}) = \begin{bmatrix} \mathbf{0}_{1\times4} & \mathbf{0}_{1\times4} & \mathbf{0}_{1\times4} & \mathbf{d}_1' \\ \mathbf{0}_{1\times4} & \mathbf{0}_{1\times4} & \mathbf{d}_2' & \mathbf{0}_{1\times4} \\ \mathbf{0}_{1\times4} & \mathbf{0}_{1\times4} & -\mathbf{d}_1' & \mathbf{d}_1' \\ \mathbf{0}_{1\times4} & \mathbf{0}_{1\times4} & \mathbf{d}_2' & -\mathbf{d}_2' \end{bmatrix} \mathrm{vec}(\mathbf{Q}) \geq \mathbf{0}_{4\times1}.$$

The identified set for the impulse response $\eta_{i,j,h}$ given the zero and sign restrictions is

$$IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F,S) = \{\eta_{i,j,h}(\boldsymbol{\phi},\mathbf{Q}) : \mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi}|F,S)\},$$

where $\mathscr{Q}(\boldsymbol{\phi}|F,S)$ is the set of orthonormal matrices that satisfy the zero and sign restrictions and the sign normalisation:

$$\mathscr{Q}(\boldsymbol{\phi}|F,S) = \{\mathbf{Q} \in \mathscr{O}(n) : \mathbf{F}(\boldsymbol{\phi},\mathbf{Q}) = \mathbf{0}_{(\sum_i^n f_i)\times1},$$
$$\mathbf{S}(\boldsymbol{\phi},\mathbf{Q}) \geq \mathbf{0}_{s\times1}, \mathrm{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times1}\}.$$

Let $\pi_{\boldsymbol{\phi}}$ be a prior over the reduced-form parameter $\boldsymbol{\phi}$. A joint prior for $\boldsymbol{\theta} = (\boldsymbol{\phi}', \mathrm{vec}(\mathbf{Q})')' \in \boldsymbol{\Phi} \times \mathrm{vec}(\mathscr{O}(n))$ can be written as $\pi_{\boldsymbol{\theta}} = \pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_{\boldsymbol{\phi}}$, where $\pi_{\mathbf{Q}|\boldsymbol{\phi}}$ is supported only on $\mathscr{Q}(\boldsymbol{\phi}|F,S)$. Under point identification, the identifying restrictions pin down a unique value of $\mathbf{Q}$ given $\boldsymbol{\phi}$. Consequently, specifying a prior for $\boldsymbol{\phi}$ is sufficient to induce a single prior – and thus a single posterior – for $\boldsymbol{\theta}$. In the set-identified case, the identifying restrictions do not uniquely determine $\mathbf{Q}$ given $\boldsymbol{\phi}$, so specifying a prior for the reduced-form parameters does not induce a single prior for $\boldsymbol{\theta}$ and thus does not yield a single posterior. Following Uhlig (2005), the vast majority of the empirical literature using Bayesian methods in set-identified SVARs imposes a single prior for $\mathbf{Q}|\boldsymbol{\phi}$, including ARW21 in their set-identified proxy SVARs. However, while the prior for $\boldsymbol{\phi}$ is updated by the data, the conditional prior for $\mathbf{Q}|\boldsymbol{\phi}$ is not updated, even asymptotically, because the likelihood does not depend on $\mathbf{Q}$ (Poirier (1998); Moon and Schorfheide (2012)). This is problematic, because posterior inference may be driven by an arbitrary prior for $\mathbf{Q}$, which has no direct economic interpretation, and even a uniform prior over $\mathcal{O}(n)$ may be informative about the objects of interest, such as impulse responses (Baumeister and Hamilton (2015)).

Rather than specifying a single prior, the robust Bayesian approach of GK considers the class of all priors for $\mathbf{Q}|\boldsymbol{\phi}$ that are consistent with the identifying restrictions:

$$\Pi_{\mathbf{Q}|\boldsymbol{\phi}} = \left\{\pi_{\mathbf{Q}|\boldsymbol{\phi}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}}(\mathscr{Q}(\boldsymbol{\phi}|F,S)) = 1\right\}.$$

Combining the class of priors with the posterior for $\boldsymbol{\phi}$ generates a class of posteriors for $\boldsymbol{\theta}$:

$$\Pi_{\boldsymbol{\theta}|\mathbf{Y},\mathbf{M}} = \left\{\pi_{\boldsymbol{\theta}|\mathbf{Y},\mathbf{M}} = \pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}} \in \Pi_{\mathbf{Q}|\boldsymbol{\phi}}\right\},$$

where $\mathbf{Y} = (\mathbf{y}'_{1-p},\ldots,\mathbf{y}'_T)'$ and $\mathbf{M} = (\mathbf{m}'_{1-p},\ldots,\mathbf{m}'_T)'$. In turn, the class of posteriors for $\boldsymbol{\theta}$ induces a class of posteriors for $\eta_{i,j,h}$. GK suggest summarising this class of posteriors by reporting the 'set of posterior means':

$$\left[\int_{\boldsymbol{\Phi}} l(\boldsymbol{\phi})d\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}, \int_{\boldsymbol{\Phi}} u(\boldsymbol{\phi})d\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}\right],$$

where $l(\boldsymbol{\phi}) = \inf\{\eta_{i,j,h}(\boldsymbol{\phi},\mathbf{Q}) : \mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi}|F,S)\}$ and $u(\boldsymbol{\phi}) = \sup\{\eta_{i,j,h}(\boldsymbol{\phi},\mathbf{Q}) : \mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi}|F,S)\}$. They also suggest reporting a robust credible region with credibility level α (see Proposition 1 of GK). This region is interpreted as the shortest interval estimate for $\eta_{i,j,h}$ such that the posterior probability put on the interval is greater than or equal to α uniformly over the posteriors in the class. One can also report posterior probability bounds, which are the lowest and highest posterior probabilities of an event over all priors in the class.

When there are zero restrictions only, the identified set is never empty and so the data are not informative about the plausibility of the identifying restrictions. When there are sign restrictions, the identified set may be empty at particular values of $\boldsymbol{\phi}$. The posterior probability that the identified set is non-empty, $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}(\{\boldsymbol{\phi} : IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F,S) \neq \emptyset\})$, can thus be used to quantify the plausibility of the identifying restrictions.

5.3 Frequentist Validity

In this section we provide conditions under which the robust Bayesian inferential approach provides valid frequentist inference about impulse responses in the proxy SVAR. This may be of interest to frequentists who use Bayesian approaches to inference purely for computational convenience. Bayesians may also be interested in the asymptotic frequentist properties of Bayesian procedures if this facilitates the communication of their results to frequentist audiences.

The set of posterior means can be interpreted as a consistent estimator of the true identified set if $IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F,S)$ is convex and is a continuous correspondence of $\boldsymbol{\phi}$ at the true value ϕ_0 (see Theorem 3 in GK). If, in addition, $l(\boldsymbol{\phi})$ and $u(\boldsymbol{\phi})$ are differentiable in $\boldsymbol{\phi}$ at $\boldsymbol{\phi}_0$ with nonzero derivatives, and the posterior for $\boldsymbol{\phi}$ satisfies the Bernstein-von Mises property, the robust credible region is an asymptotically valid confidence set for the true identified set (see Proposition 2 in GK). In the context of an SVAR, Propositions B.1 and B.2 of GK provide conditions under which the impulse-response identified set is guaranteed to be convex and continuous in $\boldsymbol{\phi}$, respectively, while Proposition B.3 provides conditions under which it is guaranteed to be differentiable in $\boldsymbol{\phi}$. In the proxy SVAR, we can show that having the relevance condition satisfied at $\boldsymbol{\phi}_0$ is a necessary condition for continuity of the identified-set correspondence at $\boldsymbol{\phi}_0$. In Section 5.3.1 we proceed under the assumption that the relevance condition is satisfied and provide conditions under which the identified-set correspondence is convex and is differentiable in $\boldsymbol{\phi}$. We explore issues associated with ‘weak’ proxies – where the relevance condition is ‘close’ to being violated – in Section 5.3.2.

5.3.1 Strong proxies

Assume that there are k proxies correlated with the last k structural shocks and uncorrelated with the remaining $n-k$ structural shocks. Assume also that $n>3$ and $1<k<n-1$, so there are multiple proxies for multiple shocks and the impulse responses to all shocks are set-identified. This is the setting in MR13 and Mertens and Montiel Olea (2018) (before the imposition of additional point-identifying zero restrictions), and so is of empirical relevance. The propositions below clarify con-

ditions for $IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F,S)$ to be convex and differentiable in $\boldsymbol{\phi}$, in which case the robust Bayesian approach provides asymptotically valid frequentist inference about the impulse-response identified set. We relegate proofs to Appendix C.1.

Proposition 5.3.1. *Let the object of interest be $\eta_{i,j^*,h} = \mathbf{c}_{i,h}(\boldsymbol{\phi})\mathbf{q}_{j^*}$, the impulse response of the ith variable at the hth horizon to the j^*th structural shock, where the variables are ordered according to Definition 1.*

(I) Suppose there are only zero restrictions arising from the exogeneity assumption and that the relevance condition holds, so $\mathrm{rank}(\mathbf{D\Sigma}_{tr}) = k$. Then, for every i and h and almost every $\boldsymbol{\phi} \in \boldsymbol{\Phi}$, the identified set of $\eta_{i,j^*,h}$, $IS_{\eta_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$, is convex.

(II) Consider the case with both zero and sign restrictions. Suppose the only zero restrictions are those arising from the exogeneity restrictions and that the relevance condition holds, so $\mathrm{rank}(\mathbf{D\Sigma}_{tr}) = k$. Also assume that any sign restrictions constrain the j^*th column of $\mathbf{Q}$ only and let $\mathbf{S}_{j^*}(\boldsymbol{\phi})\mathbf{q}_{j^*} \geq \mathbf{0}_{s\times 1}$ represent the sign restrictions.

(i) If interest is in the impulse responses to one of the first $n-k$ structural shocks, then $j^* = 1$ by Definition 1, and $IS_{\eta_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$ is convex for every i and h if there exists a unit-length vector $\mathbf{q} \in \mathbb{R}^n$ satisfying

$$\mathbf{D\Sigma}_{tr}\mathbf{q} = \mathbf{0}_{k\times 1} \quad \text{and} \quad \mathbf{S}_1(\boldsymbol{\phi})\mathbf{q} > \mathbf{0}_{s\times 1}.$$

(ii) Let $\mathbf{N}(\mathbf{D\Sigma}_{tr})$ be an orthonormal basis for the nullspace of $\mathbf{D\Sigma}_{tr}$ (so $\mathbf{N}(\mathbf{D\Sigma}_{tr})$ is an $n\times(n-k)$ matrix). If interest is in the impulse responses to one of the last k structural shocks, then $j^* = n-k+1$ by Definition 1, and $IS_{\eta_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$ and convex for every i and h if there exists a unit-length vector $\mathbf{q} \in \mathbb{R}^n$ satisfying

$$\mathbf{N}(\mathbf{D\Sigma}_{tr})'\mathbf{q} = \mathbf{0}_{k\times 1} \quad \text{and} \quad \mathbf{S}_{n-k+1}(\boldsymbol{\phi})\mathbf{q} > \mathbf{0}_{s\times 1}.$$

Proposition 5.3.1 states that, when there are exogeneity restrictions only, the identified set for the impulse response is convex for almost every $\boldsymbol{\phi} \in \boldsymbol{\Phi}$. When there are also sign restrictions constraining $\mathbf{q}_{j^*}$ only, the identified set is convex conditional on it being nonempty. Note that convexity of the identified set in the empirically relevant case, when interest is in the responses to one of the last k shocks, does not follow from Proposition B.1 of GK. The key difference from the general setting of GK is that, in our case, $\mathbf{F}_i(\boldsymbol{\phi}) = \mathbf{D}\boldsymbol{\Sigma}_{tr}$ has full row rank and is common for $i = 1, ..., j^* - 1$. These special features of the matrix of zero restrictions makes it possible to characterise the set of feasible values for $\mathbf{q}_{j^*}$.

For the same cases in which we can guarantee convexity of the impulse-response identified set, we provide sufficient conditions for the differentiability of $u(\boldsymbol{\phi})$ and $l(\boldsymbol{\phi})$. To do this, we follow GK by building on results from Gafarov *et al.* (2018), who show the directional differentiability of the upper and lower bound of the impulse-response identified set when there are zero and sign restrictions on $\mathbf{q}_{j^*}$ only.

Proposition 5.3.2. *Let the object of interest be $\eta_{i,j^*,h} = \mathbf{c}_{i,h}(\boldsymbol{\phi})\mathbf{q}_{j^*}$, where the variables are ordered according to Definition 1. Suppose the only zero restrictions are those arising from the exogeneity assumption and that the relevance condition holds, so* $\text{rank}(\mathbf{D}\boldsymbol{\Sigma}_{tr}) = k$. *Also assume that any sign restrictions constrain the j^*th column of $\mathbf{Q}$ only and let $\mathbf{S}_{j^*}(\boldsymbol{\phi})\mathbf{q}_{j^*} \geq \mathbf{0}_{s\times 1}$ represent the sign restrictions.*

(i) Suppose the impulse responses of interest are those to one of the first $n-k$ structural shocks, so $j^* = 1$ by Definition 1, and that the column vectors of $\left[(\mathbf{D}\boldsymbol{\Sigma}_{tr})', \mathbf{S}_1(\boldsymbol{\phi})', \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n}\right]$ are linearly independent at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. If, at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, the set of solutions of the optimisation problem

$$\max_{\mathbf{q}\in\mathscr{S}^{n-1}} \left(\min_{\mathbf{q}\in\mathscr{S}^{n-1}}\right) \quad \mathbf{c}_{i,h}'(\boldsymbol{\phi})\mathbf{q}$$

$$s.t. \quad \mathbf{D}\boldsymbol{\Sigma}_{tr}\mathbf{q} = \mathbf{0}_{k\times 1} \quad and \quad \begin{bmatrix} \mathbf{S}_1(\boldsymbol{\phi}) \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1)\times 1}$$

is singleton, the optimised value $u(\boldsymbol{\phi})$ ($l(\boldsymbol{\phi})$) is nonzero, and the number of binding sign restrictions at the optimum is less than or equal to $n-k-1$, then $u(\boldsymbol{\phi})$ ($l(\boldsymbol{\phi})$) is differentiable at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

(ii) Suppose the impulse responses of interest are those to one of the last k structural shocks, so $j^* = n-k+1$ by Definition 1, and that the column vectors of

$\left[\mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr}), \mathbf{S}_1(\boldsymbol{\phi})', \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n-k+1,n}\right]$ *are linearly independent at* $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. *If, at* $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, *the set of solutions of the optimisation problem*

$$\max_{\mathbf{q}\in\mathscr{S}^{n-1}} \left(\min_{\mathbf{q}\in\mathscr{S}^{n-1}}\right) \quad \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q}$$

$$s.t. \quad \mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})'\mathbf{q} = \mathbf{0}_{k\times 1} \quad and \quad \begin{bmatrix} \mathbf{S}_{n-k+1}(\boldsymbol{\phi}) \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n-k+1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1)\times 1}$$

is singleton, the optimised value $u(\boldsymbol{\phi})$ *(*$l(\boldsymbol{\phi})$*) is nonzero, and the number of binding sign restrictions at the optimum is less than or equal to* $k-1$, *then* $u(\boldsymbol{\phi})$ *(*$l(\boldsymbol{\phi})$*) is differentiable at* $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

When $n \geq 3$, it is also straightforward to show that the identified set is convex when $k = 1$ and interest is in the impulse responses to one of the first $n-1$ structural shocks (the impulse responses to the last structural shock are point-identified), or when $k = n-1$ and interest is in the impulse responses to one of the last $n-1$ shocks (the impulse responses to the first structural shock are point-identified). Differentiability in these cases is also obtained under similar conditions to those in Proposition 5.3.2.

When there are sign restrictions that constrain multiple columns of $\mathbf{Q}$, we cannot guarantee convexity of the identified set (see Example B.5 in GK), nor differentiability. Nevertheless, the set of posterior means and robust credible region can be interpreted as providing inference about the *convex hull* of the identified set.

Since $\text{FEVD}_{i,j^*,h}$ is a continuous function of $\mathbf{q}_{j^*}$, $IS_{\text{FEVD}_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$ is continuous at $\boldsymbol{\phi}_0$ and it is convex whenever $IS_{\eta_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$ is convex. If $IS_{\text{FEVD}_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$ is also differentiable in $\boldsymbol{\phi}$, we can guarantee frequentist validity of the robust Bayesian inferential procedure when applied to the FEVD in the same cases as for the impulse response. However, we are unaware of results on the differentiability of $IS_{\text{FEVD}_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$. We leave exploration of this to further work.

5.3.2 Weak proxies

In this section we investigate how weak proxies affect robust Bayesian posterior inference about set-identified impulse responses in the proxy SVAR. Our focus is on the asymptotic frequentist properties of our procedure. We consider the case where $n = 3$, $k = 1$ and the objects of interest are the impulse responses to ε_{1t}. We choose this case because it is straightforward to analytically characterise the identified set and hence to discuss the effects of weak proxies.

At a given value of $\boldsymbol{\phi} \in \boldsymbol{\Phi}$ (and ignoring the sign normalisation), the upper bound of $IS_{\eta_{i,1,h}}(\boldsymbol{\phi}|F,S)$ is the value function associated with the following optimisation problem:[12]

$$u(\boldsymbol{\phi}) = \max_{\mathbf{q}\in\mathscr{S}^{n-1}} \mathbf{c}'\mathbf{q} \quad \text{subject to} \quad \mathbf{d}'\mathbf{q} = 0,$$

where $\mathbf{c} \equiv \mathbf{c}_{i,h}(\boldsymbol{\phi})$ and $\mathbf{d}' \equiv \mathbf{D}\boldsymbol{\Sigma}_{tr}$. Applying the change of variables $\mathbf{x} = \boldsymbol{\Sigma}_{tr}\mathbf{q}$ yields the problem in Equation (2.5) of Gafarov *et al.* (2018). Using their results, the value function satisfies

$$u(\boldsymbol{\phi})^2 = \mathbf{c}'\left[\mathbf{I}_3 - \mathbf{d}\left(\mathbf{d}'\mathbf{d}\right)^{-1}\mathbf{d}'\right]\mathbf{c}. \tag{5.8}$$

Equation (5.7) implies that $\mathbb{E}(m_t\varepsilon_{3t}) = \mathbf{d}'\mathbf{q}_3 = \Psi$. The exogeneity restrictions require that $\mathbf{q}_1$ and $\mathbf{q}_2$ are orthogonal to $\mathbf{d}$. Since the columns of $\mathbf{Q}$ are orthogonal and have unit length, $\mathbf{q}_3 = \pm\mathbf{d}/\|\mathbf{d}\|$, which implies that $\mathbf{d}'\mathbf{d}/\|\mathbf{d}\| = \|\mathbf{d}\| = |\Psi|$.

A 'weak' proxy correlates with one of the structural shocks only weakly, so $|\Psi|$ is close to zero. This is equivalent to $\|\mathbf{d}\|$ being small. Note that $u(\boldsymbol{\phi})$ as the square root of (5.8) is continuous and smooth in $\mathbf{c}$, while it is discontinuous in $\mathbf{d}$ at $\mathbf{d} = \mathbf{0}_{3\times 1}$. Hence, if the posterior distribution of $\mathbf{d}$ concentrates near a point of singularity of $u(\boldsymbol{\phi})$ due to the weak proxy, the posterior of $u(\boldsymbol{\phi})$ can exhibit a nonstandard distribution even when the posterior of $(\mathbf{c},\mathbf{d})$ is consistent and can be well approximated by a normal distribution centered at the maximum likelihood estimator (MLE).

To investigate the posterior for $u(\boldsymbol{\phi})$ in the weak-proxy case, we consider the local asymptotic approximation of the posterior for $u(\boldsymbol{\phi})$ with a drifting sequence of the true values of $\boldsymbol{\phi}$ converging to a point of singularity. We here present the heuristic exposition of the results and defer the regularity conditions and formal

[12] In the absence of sign normalisation restrictions, the lower bound of the identified set $l(\boldsymbol{\phi})$ is given by $-u(\boldsymbol{\phi})$. This section hence focuses only on the posterior for $u(\boldsymbol{\phi})$.

proofs to Appendix C.2.

We consider a drifting sequence of data-generating processes $\{\boldsymbol{\phi}_T : T = 1,2,\ldots\}$ that induces a drifting sequence of parameter values $\{(\mathbf{c}_T, \mathbf{d}_T) : T = 1,2\ldots,\}$ converging to a point of singularity. Following the weak-instrument asymptotics of Staiger and Stock (1997), we consider the drifting sequence of $(\mathbf{c}, \mathbf{d})$ with $T^{-1/2}$-convergence rate,

$$\mathbf{c}_T = \mathbf{c}_0 + \frac{\boldsymbol{\gamma}}{\sqrt{T}}, \quad \mathbf{d}_T = \frac{\boldsymbol{\delta}}{\sqrt{T}}, \tag{5.9}$$

where $\mathbf{c}_0 \neq \mathbf{0}_{3\times 1}$, $\boldsymbol{\delta} \neq \mathbf{0}_{3\times 1}$ and $(\boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^3 \times \mathbb{R}^3$ are the localisation parameters. The magnitude of $\boldsymbol{\delta}$ characterises the relevance of the proxy; that is, a smaller value of $\|\boldsymbol{\delta}\|$ implies a weaker proxy.

Let $(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$ be the MLE for $(\mathbf{c}, \mathbf{d})$ (which is a constant once we have conditioned on the sample). We assume that the sampling distribution of the MLE is $\sqrt{T}$-asymptotically normal:

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \hat{\mathbf{c}}_T - \mathbf{c}_T \\ \hat{\mathbf{d}}_T - \mathbf{d}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}_{6\times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}_{cd}' & \boldsymbol{\Omega}_d \end{pmatrix} \right). \tag{5.10}$$

We also assume that the posterior for $(\mathbf{c}, \mathbf{d})$ converges to a normal distribution with data-independent variance. That is, conditional on the sampling sequence,

$$\sqrt{T} \begin{pmatrix} \mathbf{c} - \hat{\mathbf{c}}_T \\ \mathbf{d} - \hat{\mathbf{d}}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{Z}_c \\ \mathbf{Z}_d \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}_{6\times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}_{cd}' & \boldsymbol{\Omega}_d \end{pmatrix} \right), \tag{5.11}$$

as $T \to \infty$ for almost every sampling sequence, where $\boldsymbol{\Omega} \equiv \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}_{cd}' & \boldsymbol{\Omega}_d \end{pmatrix}$ is the posterior asymptotic variance, which does not depend on the sampling sequence. The asymptotic equivalence of the probability laws in (5.10) and (5.11) implies that the reduced-form parameters $(\mathbf{c}, \mathbf{d})$ are regular in the sense that the well-known Bernstein-von Mises Theorem holds. See, for instance, Schervish (1995) and DasGupta (2008) for a set of sufficient conditions for posterior asymptotic normality with the Bernstein-von Mises property.

Under this setting, Proposition C.2.1 in Appendix C.2 derives the following asymptotic approximation of the posterior for $u(\boldsymbol{\phi})$. Conditional on the sampling sequence,

$$u(\boldsymbol{\phi}) \xrightarrow{d} \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d\|^2} \right) \mathbf{c}_0}, \tag{5.12}$$

as $T \to \infty$ for almost every sampling sequence, where $\hat{\mathbf{Z}}_d$ is a constant that depends on the sample, and $\mathbf{Z}_d \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Omega}_d)$.

This representation of the asymptotic posterior provides the following insights about the influence of the weak proxy on posterior inference. First, the posterior of $u(\boldsymbol{\phi})$ is not consistent and remains a non-degenerate distribution in large samples. Second, the asymptotic posterior for $u(\boldsymbol{\phi})$ depends not only on the localisation parameter $\boldsymbol{\delta}$, but also on the statistic $\hat{\mathbf{Z}}_d$ realized in the data. Hence, unlike in the well-identified case, the influence of the data on the shape of the posterior does not disappear in large samples. Also, the asymptotic posterior mean almost always (in terms of the sampling probability) misses the upper bound of the true identified set defined by the limit along the drifting data-generating processes $\{\boldsymbol{\phi}_T : T = 1, 2, \dots\}$ yielding (5.9):

$$\lim_{T\to\infty} u(\boldsymbol{\phi}_T) = \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{\|\boldsymbol{\delta}\|^2} \right) \mathbf{c}_0}.$$

This implies that, under the current weak-proxy asymptotics, the set of posterior means for the impulse response is not a consistent estimator for the identified set.

Under the same drifting sequence inducing (5.9), Proposition C.2.2 in Appendix C.2 derives the asymptotic sampling distribution of the MLE for the upper bound of the identified set:

$$u(\hat{\boldsymbol{\phi}}) \xrightarrow{d} \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d\|^2} \right) \mathbf{c}_0}, \tag{5.13}$$

where $\hat{\mathbf{Z}}_d \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Omega}_d)$. Like the posterior for $u(\boldsymbol{\phi})$, the sampling distribution of $u(\hat{\boldsymbol{\phi}})$ is not consistent and remains non-degenerate. A comparison of (5.12) and

(5.13) shows that the posterior and the sampling distribution of the MLE for the bound of the identified set do not asymptotically coincide for almost every sampling sequence. The Bernstein-von Mises property therefore fails for the estimation of the upper and lower bounds of the identified set. This implies that the asymptotic frequentist coverage of the robust credible region of GK can also fail, because a condition analogous to Assumption 4 in GK 18 does not hold in the current setting of weak proxy asymptotics.

When the proxy is strong in the sense that $|\Psi| = \|\mathbf{d}\|$ is far from zero, the pointwise asymptotic approximation of the posterior of $u(\boldsymbol{\phi})$ approximates well the finite-sample posterior. Noting that $u(\boldsymbol{\phi})$ is smooth at $\mathbf{d} \neq \mathbf{0}_{3\times 1}$ and assuming that the posterior of $(\mathbf{c}, \mathbf{d})$ centered at the MLE is $\sqrt{T}$-asymptotically normal, the delta method implies that $\sqrt{T}(u(\boldsymbol{\phi}) - u(\hat{\boldsymbol{\phi}}))$ is asymptotically normal with a data-independent variance. This asymptotic posterior coincides with the sampling distribution of the MLE, so correct frequentist coverage of the robust credible region can be attained in addition to posterior consistency. This stark contrast in the asymptotic behavior of the posteriors suggests that, in the current simple setting, whether the posterior of $u(\boldsymbol{\phi})$ is non-normal could be useful for diagnosing whether the proxy is weak. We leave a formal analysis of this for future research.

5.4 Numerical Implementation

In this section, we provide numerical algorithms to conduct robust Bayesian inference about set-identified objects of interest in proxy SVARs. The algorithms numerically approximate the set of posterior means and associated robust credible interval. When there are sign restrictions, the algorithms also give estimates of the plausibility of the identifying restrictions. Throughout, we assume that the order of the variables satisfies Definition 1. Since the identifying restrictions are linear restrictions on columns of $\mathbf{Q}$, the algorithms are similar to the algorithms in GK. We repeat them here for completeness and discuss further details, and issues specific to the proxy SVAR case, below. Algorithm 1 assumes that the object of interest is the impulse response; the subsequent remarks discuss how to conduct robust Bayesian inference about other objects of interest. Matlab code implementing the algorithms is available on the authors' personal websites.

Algorithm 1. *Let* $\mathbf{F}(\boldsymbol{\phi},\mathbf{Q}) = \mathbf{0}_{(\sum_{i=1}^{n} f_i)\times 1}$ *and* $S(\boldsymbol{\phi},\mathbf{Q}) \geq \mathbf{0}_{s\times 1}$ *be the set of identifying restrictions and let* $\eta_{i,j^*,h} = \mathbf{c}'_{i,h}\mathbf{q}_{j^*}$ *be the impulse response of interest.*

- ***Step 1**: Specify a prior for* $\boldsymbol{\phi}$, $\pi_{\boldsymbol{\phi}}$, *and obtain the posterior* $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}$.[13]

- ***Step 2**: Draw* $\boldsymbol{\phi}$ *from* $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}$ *and check whether* $\mathscr{Q}(\boldsymbol{\phi}|F,S)$ *is empty using the subroutine below.*

  - ***Step 2.1**: Draw* $\mathbf{z}_1 \sim N(\mathbf{0}_{n\times 1},\mathbf{I}_n)$ *and let* $\tilde{\mathbf{q}}_1 = \left[\mathbf{I}_n - \mathbf{F}'_1(\mathbf{F}_1\mathbf{F}'_1)^{-1}\mathbf{F}_1\right]\mathbf{z}_1$. *For* $i = 2,\ldots,n$, *run the following procedure sequentially: draw* $\mathbf{z}_i \sim N(\mathbf{0}_{n\times 1},\mathbf{I}_n)$ *and compute* $\tilde{\mathbf{q}}_i = \left[\mathbf{I}_n - \tilde{\mathbf{F}}'_i(\tilde{\mathbf{F}}_i\tilde{\mathbf{F}}'_i)^{-1}\tilde{\mathbf{F}}_i\right]\mathbf{z}_i$, *where* $\tilde{\mathbf{F}}'_i = [\mathbf{F}'_i,\tilde{\mathbf{q}}_1,\ldots,\tilde{\mathbf{q}}_{i-1}]$.

  - ***Step 2.2**: Given* $\tilde{\mathbf{q}}_i$, $i = 1,\ldots,n$, *define*

$$\mathbf{Q}_0 = \left[\operatorname{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\tilde{\mathbf{q}}_1)\frac{\tilde{\mathbf{q}}_1}{\|\tilde{\mathbf{q}}_1\|},\ldots,\operatorname{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n,n})'\tilde{\mathbf{q}}_n)\frac{\tilde{\mathbf{q}}_n}{\|\tilde{\mathbf{q}}_n\|}\right].$$ [14]

  - ***Step 2.3**: Check whether* $\mathbf{Q}_0$ *satisfies* $\mathbf{S}(\boldsymbol{\phi},\mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$. *If so, retain* $\mathbf{Q}_0$ *and proceed to Step 3. Otherwise, repeat Steps 2.1 and 2.2 (up to a maximum of L times) until* $\mathbf{Q}_0$ *is obtained satisfying* $\mathbf{S}(\boldsymbol{\phi},\mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$. *If no draws of* $\mathbf{Q}_0$ *satisfy* $\mathbf{S}(\boldsymbol{\phi},\mathbf{Q}_0) \geq \mathbf{0}_{s\times 1}$, *approximate* $\mathscr{Q}(\boldsymbol{\phi}|F,S)$ *as being empty and return to Step 2.*

- ***Step 3**: Compute the lower bound of* $IS_{\eta_{i,j^*,h}}(\boldsymbol{\phi}|F,S)$ *by solving the following constrained optimisation problem with initial value* $\mathbf{Q}_0$*:*

$$l(\boldsymbol{\phi}) = \min_{\mathbf{Q}} \mathbf{c}'_{i,h}\mathbf{q}_{j^*}$$

[13] $\pi_{\boldsymbol{\phi}}$ does not have to be proper or to satisfy the condition $\pi_{\boldsymbol{\phi}}(\{\boldsymbol{\phi} : \mathscr{Q}(\boldsymbol{\phi}|F,S) \neq \emptyset\}) = 1$ for all $\boldsymbol{\phi} \in \boldsymbol{\Phi}$; that is, $\pi_{\boldsymbol{\phi}}$ may assign positive probability to regions of $\boldsymbol{\Phi}$ that yield an empty set of orthonormal matrices satisfying the identifying restrictions.

[14] If $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{i,n})'\tilde{\mathbf{q}}_i = 0$ for some i, set $\operatorname{sign}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\tilde{\mathbf{q}}_i)$ equal to 1 or -1 with equal probability.

subject to

$$\begin{aligned} \mathbf{F}(\boldsymbol{\phi},\mathbf{Q}) &= \mathbf{0}_{(\sum_i^n f_i)\times 1} \\ \mathbf{S}(\boldsymbol{\phi},\mathbf{Q}) &\geq \mathbf{0}_{s\times 1} \\ \mathrm{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) &\geq \mathbf{0}_{n\times 1} \\ \mathbf{Q}'\mathbf{Q} &= \mathbf{I}_n. \end{aligned}$$

Similarly, obtain $u(\boldsymbol{\phi}) = \max_{\mathbf{Q}} \mathbf{c}_{i,h}'\mathbf{q}_{j^*}$ *under the same set of constraints.*

- ***Step 4**: Repeat Steps 2–3 M times to obtain* $[l(\boldsymbol{\phi}_m), u(\boldsymbol{\phi}_m)]$ *for* $m = 1,...,M$. *Approximate the set of posterior means by the sample averages of* $l(\boldsymbol{\phi}_m)$ *and* $u(\boldsymbol{\phi}_m)$.

- ***Step 5**: To obtain an approximation of the smallest robust credible region with credibility* $\alpha \in (0,1)$, *define* $d(\eta,\boldsymbol{\phi}) = \max\{|\eta - l(\boldsymbol{\phi})|, |\eta - u(\boldsymbol{\phi})|\}$ *and let* $\hat{z}_\alpha(\eta)$ *be the sample* α*-th quantile of* $\{d(\eta,\boldsymbol{\phi}_m), m = 1,...,M\}$. *An approximated smallest robust credible interval for* $\eta_{i,j^*,h}$ *is an interval centered at* $\arg\min_\eta \hat{z}_\alpha(\eta)$ *with radius* $\min_\eta \hat{z}_\alpha(\eta)$.

- ***Step 6**: Approximate* $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}(\{\boldsymbol{\phi} : \mathscr{Q}(\boldsymbol{\phi}|F,S) \neq \emptyset\})$ *by the proportion of draws of* $\boldsymbol{\phi}$ *passing Step 2.3.*

5.4.1 Remarks

5.4.1.1 Further details about Step 2

Given a draw of $\boldsymbol{\phi}$ from its posterior, Step 2 attempts to draw $\mathbf{Q}$ satisfying the zero and sign restrictions. The vectors $\tilde{\mathbf{q}}_i$, $i = 1,\ldots,n$, are residual vectors from the linear projection of multivariate standard normally distributed random variables on vectors representing the zero restrictions and previously constructed columns of $\mathbf{Q}$. These residual vectors therefore satisfy the zero restrictions represented in $\mathbf{F}(\boldsymbol{\phi},\mathbf{Q})$ and are orthogonal.[15] Step 2.2 rescales the residual vectors to have unit length and

[15] If the relevance condition fails, $\mathbf{F}_i(\boldsymbol{\phi})$ is of reduced row rank for $i = 1,\ldots,n-k$ and the coefficients in the linear projection are not identified. This is a measure zero event so long as $\pi_{\boldsymbol{\phi}}$ does not place positive probability mass on the event $\mathrm{rank}(\mathbf{D}) < k$.

imposes the sign normalisation that the diagonal elements of $\mathbf{A}_0$ are nonnegative.

Step 2 can be interpreted as implementing a particular implicit prior over $\mathscr{Q}(\boldsymbol{\phi}|F,S)$. This implicit prior is irrelevant for the class of posteriors generated by the robust Bayesian procedure, since the draw of $\mathbf{Q}$ is used only as an initial value in the numerical optimisation step. However, one could use these draws to construct the posterior for $\eta_{i,j,h}$ induced by this prior; in the empirical application below, we do this to illustrate how posterior inference may be sensitive to the choice of prior for $\mathbf{Q}|\boldsymbol{\phi}$. The algorithm used to draw $\mathbf{Q}$ possesses similarities to the algorithms described in Arias et al. (2018), who conduct single-prior Bayesian inference in set-identified SVARs. However, differences in the algorithms mean that the implicit priors are different. Algorithm 3 in Arias *et al.* (2018) draws a value of $\mathbf{Q}$ satisfying the zero restrictions only once at each draw of $\boldsymbol{\phi}$ and discards the joint draw of $\boldsymbol{\phi}$ and $\mathbf{Q}$ if the sign restrictions are not satisfied. In contrast, we draw $\mathbf{Q}$ satisfying the zero restrictions until we obtain a value satisfying the sign restrictions. Relative to the prior implicit in the second approach, the prior implicit in the first approach places more weight on values of $\boldsymbol{\phi}$ with a larger identified set for $\mathbf{Q}$ (see Uhlig (2017) for a discussion of this point).

5.4.1.2 Choice of prior

The method used to draw $\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}$ depends on the posterior, and thus on the prior. In the empirical application below we use independent (improper) Jeffreys' priors over the blocks of reduced-form parameters in the VAR for $\mathbf{y}_t$ and the first-stage regression; that is, $\pi_{\boldsymbol{\phi}} = \pi_{\mathbf{B},\boldsymbol{\Sigma}}\pi_{\mathbf{D},\mathbf{J},\boldsymbol{\Upsilon}}$, where $\pi_{\mathbf{B},\boldsymbol{\Sigma}} \propto |\boldsymbol{\Sigma}|^{-\frac{n+1}{2}}$ and $\pi_{\mathbf{D},\mathbf{J},\boldsymbol{\Upsilon}} \propto |\boldsymbol{\Upsilon}|^{-\frac{k+1}{2}}$.[16] This makes it simple to draw from the posterior of $\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}$, since it is the product of independent normal-inverse-Wishart posteriors.[17] We emphasise that our algorithm does not rely on using independent priors over the reduced-form parameters; all that matters is that one can sample from the posterior of $\boldsymbol{\phi}$. For example, if the prior

[16] $\pi_{\mathbf{B},\boldsymbol{\Sigma}}$ is nonzero only for values of $\mathbf{B}$ such that the VAR is invertible into a VMA(∞).

[17] This follows from the fact that the joint likelihood of $(\mathbf{M},\mathbf{Y})$ is multiplicatively separable across the two blocks of parameters: $\pi_{\mathbf{Y},\mathbf{M}|\boldsymbol{\phi}} = \pi_{\mathbf{M}|\mathbf{Y},\mathbf{D},\mathbf{J},\boldsymbol{\Upsilon}}\pi_{\mathbf{Y}|\mathbf{B},\boldsymbol{\Sigma}}$. For an algorithm that draws from the normal-inverse-Wishart posterior distribution, see Del Negro and Schorfheide (2011). Imposing independent normal-inverse-Wishart priors would also yield a posterior that is the product of independent normal-inverse-Wishart posteriors.

is over the model's structural – rather than reduced-form – parameters, one could draw from the posterior of the structural parameters and transform these draws into draws of the reduced-form parameters.

5.4.1.3 Convergence issues and alternative algorithms

The optimisation problem in Step 3 is nonconvex. Consequently, the convergence of gradient-based optimisation methods in this problem is not guaranteed. Accordingly, we suggest drawing multiple values of $\mathbf{Q}_0$ in Steps 2.1–2.3 to use as initial values in the optimisation step, and computing optima over the set of solutions obtained from the different initial values. GK also provide an algorithm that can be used to check for convergence of, or as an alternative to, the numerical optimisation step.

Algorithm 2. *In Algorithm 1, replace Step 3 with the following:*

- ***Step 3'**: Iterate Steps 2.1–2.3 K times and let* $\{\mathbf{Q}_l, l = 1, ..., \tilde{K}\}$ *be the* $\tilde{K}$ *draws of* $\mathbf{Q}$ *that satisfy the identifying restrictions. Let* $\mathbf{q}_{j^*,l}$ *be the* j^**th column of* $\mathbf{Q}_l$. *Approximate* $[l(\boldsymbol{\phi}), u(\boldsymbol{\phi})]$ *by* $[\min_l \mathbf{c}'_{i,h}\mathbf{q}_{j^*,l}, \max_l \mathbf{c}'_{i,h}\mathbf{q}_{j^*,l}]$.

Algorithm 2 yields an approximated identified set that is smaller than the true identified set at every draw of $\boldsymbol{\phi}$. However, the approximated identified set will converge to the true identified set as $\tilde{K}$ goes to infinity. In our implementation of this algorithm, we fix $\tilde{K}$ and let K vary. We suggest determining an appropriate value of $\tilde{K}$ by fixing the value of $\boldsymbol{\phi}$ (e.g., at the MLE) and comparing the bounds obtained given different values of $\tilde{K}$. In some cases, Algorithm 2 may be computationally less demanding than Algorithm 1. For example, when the dimension of the VAR is large or if interest is in the impulse responses of many variables at many horizons, the computational cost of generating a sufficiently large number of draws of $\mathbf{Q}$ to accurately approximate the bounds of the identified sets may be smaller than the cost of carrying out the optimisation step for each variable of interest at each horizon (particularly when using multiple initial values). Conversely, Algorithm 2 may be computationally more demanding when there are sign restrictions that substantially truncate the support of $\mathbf{Q}$, because many draws of $\mathbf{Q}$ will be rejected. In practice, Step 3 and Step 3' are parallelisable, so large reductions in computing time are possible in both algorithms by distributing computation across multiple processors.

Under constraints on $\mathbf{q}_{j^*}$ only, Gafarov *et al.* (2018) develop an algorithm to compute the bounds of the identified set using an analytical expression for the bounds given a set of active zero and/or sign restrictions. This approach will typically be computationally more efficient than approximating the bounds via gradient-based numerical optimisation or simulation. However, it is not generally applicable in proxy SVARs that are likely to be of interest empirically, because the types of sign restrictions on $\mathbf{\Psi}$ that naturally arise in this setting will usually constrain multiple columns of $\mathbf{Q}$.[18]

5.4.1.4 Point identification

If $f_{j^*} = j^* - 1$, the equality restrictions on $\mathbf{q}_{j^*}$ are sufficient to point identify the object of interest. This means that the prior for $\boldsymbol{\phi}$ induces a single posterior for the object of interest. In this case, Steps 1 and 2.1–2.2 of Algorithm 1 can still be used to draw from this posterior. Because $\mathbf{q}_{j^*}$ is exactly identified, any draw of $\mathbf{Q}$ satisfying the zero restrictions will contain the same $\mathbf{q}_{j^*}$ and thus will yield the same object of interest. We make use of this in the empirical application below when estimating a proxy SVAR under point-identifying restrictions.

5.4.1.5 Other objects of interest

When interest is in the FEVD rather than the impulse response, Algorithms 1 and 2 can be modified by replacing $\mathbf{c}_{i,h}(\boldsymbol{\phi})'\mathbf{q}_{j^*}$ with $\text{FEVD}_{i,j^*,h}$. If one is interested in both impulse responses and FEVDs, Algorithm 2 may deliver large gains in computation time over Algorithm 1, because the same draws of $\mathbf{Q}$ can be used to compute bounds for all objects of interest rather than having to carry out the numerical optimisation

[18] There are special cases where the results in Gafarov *et al.* (2018) could be extended to compute the bounds of the identified set. For example, if $n > 3$, $1 < k < n-1$ and $j^* = n-k+1$, one could compute an orthonormal basis for the nullspace of $\mathbf{D\Sigma}_{tr}$ and include the vectors representing this basis in the set of active restrictions. If $k = n-1$ and $j^* = 2$, one could include the restriction $\mathbf{q}_1'\mathbf{q}_2 = 0$ in the set of active restrictions, since in this case $\mathbf{q}_1$ is point-identified. As in the example in Section 5.3.2, the analytical results could also be applied if there are exogeneity restrictions only and interest is in the (set-identified) impulse response to the first structural shock. These examples would all still require that any sign restrictions constrain only $\mathbf{q}_{j^*}$.

step for each object separately. Note also that when interest is in the *cumulative* impulse response, $\mathbf{c}_{i,h}(\boldsymbol{\phi})'\mathbf{q}_{j^*}$ is replaced with $\left(\sum_{k=1}^{h}\mathbf{c}_{i,k}(\boldsymbol{\phi})'\right)\mathbf{q}_{j^*}$.

5.4.1.6 Impulse responses to a unit shock

The algorithms above impose the normalisation $\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{I}_n$, which is typical in set-identified SVARs (e.g., Uhlig (2005)). This means that the impulse responses are to a standard-deviation shock. Algorithm 3 shows how to obtain the set of posterior means and the robust credible interval for impulse responses to a unit shock, which may be of more interest in particular applications (see Stock and Watson (2016, 2018) for a discussion of this point).

Algorithm 3. *In Algorithm 1, replace Step 3 with the following:*

- ***Step 3":*** *Iterate Steps 2.1–2.3 K times and let* $\{\mathbf{Q}_l, l = 1,...,\tilde{K}\}$ *be the* $\tilde{K}$ *draws of* $\mathbf{Q}$ *that satisfy the identifying restrictions. Let* $\mathbf{A}_{0,l}^{-1} = \boldsymbol{\Sigma}_{tr}\mathbf{Q}_l$ *and compute* $\mathbf{a}_{j^*,l} = (\mathbf{A}_{0,l}^{-1}\mathbf{e}_{j^*,n})/(\mathbf{e}_{j^*,n}'\mathbf{A}_{0,l}^{-1}\mathbf{e}_{j^*,n})$. *Approximate* $[l(\boldsymbol{\phi}), u(\boldsymbol{\phi})]$ *by* $[\min_l \mathbf{e}_{i,n}'\mathbf{C}_h\mathbf{a}_{j^*,l}, \max_l \mathbf{e}_{i,n}'\mathbf{C}_h\mathbf{a}_{j^*,l}]$.

The algorithm generates impulse responses to a standard-deviation shock that are consistent with the identifying restrictions, rescales the impulse responses so that they are with respect to a unit shock (the ith element of $\mathbf{a}_{j^*,l}$ is equal to one), and computes the bounds of the identified set using the extreme values of the rescaled impulse responses. One potential issue is that the set of posterior means and robust credible interval may be unbounded when the relevant diagonal elements of $\mathbf{A}_0^{-1} = \boldsymbol{\Sigma}_{tr}\mathbf{Q}$ are not bounded away from zero for all $\boldsymbol{\phi} \in \boldsymbol{\Phi}$ and $\mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi}|F,S)$.

5.5 Empirical Application

We illustrate our methodology using the proxy SVAR considered in MR13, who estimate the macroeconomic effects of shocks to average personal and corporate income tax rates in the United States. The variables included in their benchmark specification are the average personal income tax rate (APITR), the average corporate income tax rate (ACITR), the personal income tax base, the corporate income tax base, government purchases of final goods, gross domestic product and federal government debt. The last five variables are in real per capita terms and are included in logs. MR13 decompose the sequence of plausibly exogenous changes in tax liabilities constructed by Romer and Romer (2010) into those related to personal income taxes and those related to corporate income taxes, and they exclude changes in tax liabilities with a lag between announcement and implementation of more than one quarter. These changes in tax liabilities are divided by the relevant tax base in the previous quarter and the resulting variables are used as proxies for structural shocks to the APITR and ACITR. The data are quarterly and run from 1950Q1 to 2006Q4. The VAR includes a constant and four lags of the endogenous variables. See MR13 for further details about the construction of the variables used in the VAR and the proxies.[19]

When the objects of interest are impulse responses to $\varepsilon_{APITR,t}$, any ordering of the variables such that $\mathbf{y}_t = [\mathbf{x}_t, APITR_t, ACITR_t]$, where $\mathbf{x}_t$ contains all variables other than $APITR_t$ and $ACITR_t$, will satisfy Definition 1. When interest is in the impulse responses to $\varepsilon_{ACITR,t}$, any ordering such that $\mathbf{y}_t = [\mathbf{x}_t, ACITR_t, APITR_t]$ will satisfy Definition 1. In both cases, $f_i = 2$ for $i = 1, \ldots, 5$, $f_6 = f_7 = 0$ and $j^* = 6$. Let $\mathbf{m}_t = (m_{APITR,t}, m_{ACITR,t})'$, where $m_{APITR,t}$ and $m_{ACITR,t}$ are the rescaled changes in personal and corporate income tax liabilities, respectively. MR13 impose the identifying restrictions that $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{(1:5),t}) = \mathbf{0}_{2\times 5}$ and $\mathbb{E}(\mathbf{m}_t \boldsymbol{\varepsilon}'_{6:7}) = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is an (unknown) full-rank 2×2 matrix. These identifying restrictions are insufficient to point identify any structural shock. As discussed in MR13, if one were willing to assume that $m_{APITR,t}$ is uncorrelated with $\varepsilon_{ACITR,t}$, or vice versa for $m_{ACITR,t}$ and $\varepsilon_{APITR,t}$, the additional zero restriction would be sufficient to point identify both structural shocks of interest.[20] However, positive correlation between the proxies

[19] We obtained the data from Karel Mertens' website: https://karelmertens.com/research/.

[20] $\mathbf{q}_i$, $i = 1, \ldots, 5$, is restricted to the 5-dimensional subspace of $\mathbb{R}^7$ in the nullspace of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$. $\mathbf{q}_6$ and $\mathbf{q}_7$ therefore lie in the 2-dimensional subspace spanned by the rows of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$. If interest is in impulse responses to $\varepsilon_{APITR,t}$ and $\mathbb{E}(m_{ACITR,t}\varepsilon_{APITR,t}) = \mathbf{d}_2'\mathbf{q}_6 = 0$, $\mathbf{q}_6$ is additionally constrained to be orthogonal to $\mathbf{d}_2$ and so lies in a 1-dimensional subspace. $\mathbf{q}_7$ is orthogonal to $\mathbf{q}_6$, and so lies in the 1-dimensional subspace spanned by $\mathbf{d}_2$. Assuming that $\mathbb{E}(m_{APITR,t}\varepsilon_{ACITR,t}) = 0$ yields point identification through similar reasoning (given a re-ordering of the variables to satisfy Definition 1). Assuming that $\mathbb{E}(m_{APITR,t}\varepsilon_{ACITR,t}) = 0$ and $\mathbb{E}(m_{ACITR,t}\varepsilon_{APITR,t}) = 0$ would yield one overidentifying restriction, but our algorithms do not allow for this.

suggests that these assumptions may be inappropriate.

To achieve point identification, MR13 consider additional zero restrictions on the direct contemporaneous response of one tax rate to the other (i.e., causal orderings). For example, when interest is in the impulse responses to $\varepsilon_{APITR,t}$, they assume that the ACITR does not respond directly to a structural shock in the APITR on impact. In our setting, this restriction is $\mathbf{e}_{7,7}'\mathbf{A}_0\mathbf{e}_{6,7} = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{7,7})'\mathbf{q}_6 = 0$. This restriction also point-identifies the ACITR shock. To assess robustness of their results, they consider the alternative causal ordering that the APITR does not respond directly to a structural shock in the ACITR on impact. Either of these zero restrictions could be violated if, for instance, there are constraints that impinge on the ability of the government to change personal and corporate income tax rates independently of one another. Accordingly, we extend their robustness analysis by providing an estimator of the set of impulse responses compatible with relaxing the additional zero restriction and replacing it with a set of – arguably weaker – sign restrictions.

We assume that each proxy is positively correlated with its associated structural shock (i.e., $\mathbb{E}(m_{APITR,t}\varepsilon_{APITR,t}) \geq 0$ and $\mathbb{E}(m_{ACITR,t}\varepsilon_{ACITR,t}) \geq 0$) and that each proxy is more highly correlated with its associated structural shock than with the structural shock to the other average tax rate (i.e., $\mathbb{E}(m_{APITR,t}\varepsilon_{APITR,t}) \geq \mathbb{E}(m_{APITR,t}\varepsilon_{ACITR,t})$ and $\mathbb{E}(m_{ACITR,t}\varepsilon_{ACITR,t}) \geq \mathbb{E}(m_{ACITR,t}\varepsilon_{APITR,t})$). We also assume that the response of each average tax rate to its own structural shock is nonnegative on impact, which is a sign restriction on impulse responses (as in Uhlig (2005)). Importantly, our approach allows us to relax the additional point-identifying zero restriction while avoiding the need to impose an unrevisable prior over the model's set-identified parameters.

First, we obtain impulse responses under point-identifying restrictions. When interest is in responses to the APITR, the additional point-identifying restriction is that the ACITR does not respond directly on impact to a shock in the APITR, and vice versa when interest is in responses to the ACITR. We compare estimates under these restrictions against those obtained under the set-identifying restrictions and using the single prior for $\mathbf{Q}|\boldsymbol{\phi}$ implied by Steps 2.1–2.3 of Algorithm 1 (see the discussion in Section 5.4). The purpose of this exercise is to explore the effect of the additional zero restriction on posterior inference. We then compare the impulse responses under the set-identifying restrictions and the single prior against those obtained using our robust Bayesian approach. This isolates the effect of the single prior on posterior inference. To quantify the sensitivity of posterior inference in this model to the choice of prior for $\mathbf{Q}$, we report the 'prior informativeness' statistic proposed in GK, which measures the extent to which the Bayesian credible region is tightened by choosing a particular prior:

$$\text{Prior informativeness} = \\ 1 - \frac{\text{Width of Bayesian credible region for } \eta_{i,j,h} \text{ with credibility } \alpha}{\text{Width of robust Bayesian credible region for } \eta_{i,j,h} \text{ with credibility } \alpha}.$$

As discussed in Section 5.4, we assume independent Jeffreys' priors over the reduced-form parameters such that the VAR for $\mathbf{y}_t$ is invertible into a VMA(∞). The posterior is the product of independent normal-inverse-Wishart distributions, from which it is straightforward to obtain independent draws. We obtain 10,000 draws from the posterior of $\boldsymbol{\phi}$ with non-empty identified set. In the first-stage regression, we include a constant and exclude lags of the proxies. In this application, the optimisation step of Algorithm 1 is slow due to the dimension of the VAR and the number of horizons considered. Consequently, we use Algorithm 2 with $\tilde{K} = 10,000$ to approximate the bounds of the identified set at each draw of $\boldsymbol{\phi}$ via simulation.[21] If we cannot obtain a single draw of $\mathbf{Q}$ satisfying the sign restrictions after 100,000 draws satisfying the zero restrictions, we approximate the identified set as being empty at that draw of $\boldsymbol{\phi}$. Under the additional zero restriction in MR13, we obtain the point-identified object of interest at each draw of $\boldsymbol{\phi}$ by drawing a single value of $\mathbf{Q}$ using Steps 2.1–2.2 of Algorithm 1.

Figure 5.1 plots impulse responses to a positive standard-deviation shock in the APITR under the point-identifying restrictions, under the set-identifying restrictions with a single prior for $\mathbf{Q}|\boldsymbol{\phi}$, and under the set-identifying restrictions with the ro-

[21] We have verified that using 10,000 draws of $\mathbf{Q}$ is sufficient to accurately approximate the bounds of the identified set for the impulse response and FEVD at the MLE of $\boldsymbol{\phi}$.

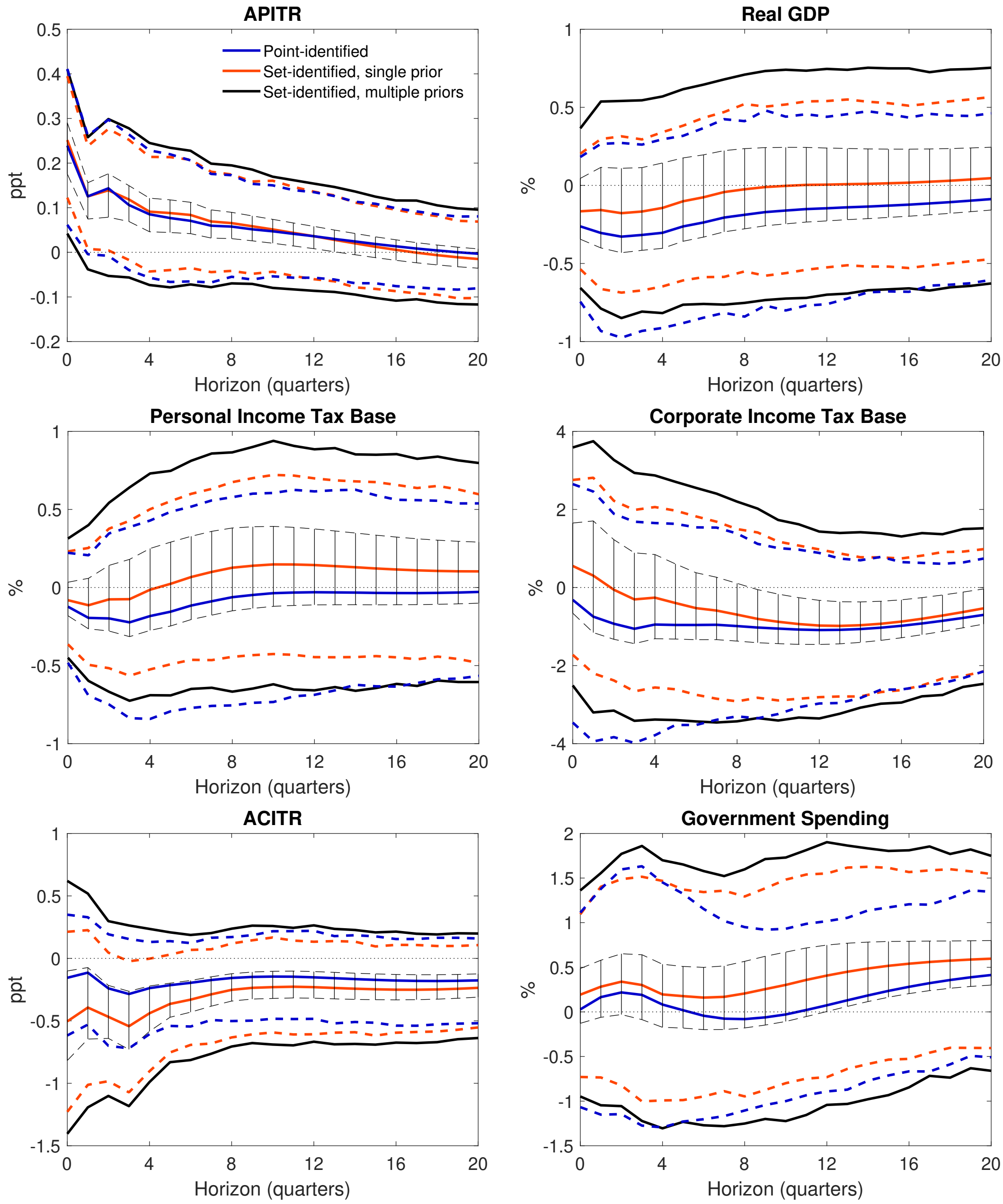
bust Bayesian approach.[22] The posterior distribution of the response of the APITR is similar under the point- and set-identifying restrictions when the single prior is used. Focusing on the output response, the 90 per cent highest posterior density (HPD) credible intervals include zero at all horizons under both sets of restrictions. Considering the class of all priors consistent with the set-identifying restrictions widens the credible intervals further; the prior informativeness statistic indicates that the choice of the single prior shrinks the width of the 90 credible interval for the output response by about 25 per cent on average over the horizons considered.

Figure 5.2 repeats Figure 5.1 for a shock to the ACITR. The response of the ACITR to its own shock is qualitatively similar under the two sets of identifying restrictions when a single prior is used. Under the point-identifying restrictions, the 90 per cent HPD intervals for the output response include zero at all horizons, which suggests that shocks to the AICTR have no effect on output. In contrast, under the set-identifying restrictions, the HPD intervals exclude zero at short horizons. However, inferences about the response of output are sensitive to the choice of single prior; the 90 per cent robust credible intervals for the output response include zero at all horizons and the prior informativeness statistic is about 20 per cent on average over the horizons considered.

Figure 5.3 plots the FEVD of output with respect to the two income tax shocks. Focusing on the posterior mean of the FEVD, the APITR shock accounts for about 20 per cent of the forecast error variance at the one-year horizon under the point-identifying restrictions. This figure falls to 10 per cent under the set-identifying restrictions and the single prior, but the result is sensitive to the choice of prior for $\mathbf{Q}|\boldsymbol{\phi}$; the set of posterior means includes values from about 5 per cent to about 25 per cent. Under the point-identifying restrictions, the ACITR shock accounts for

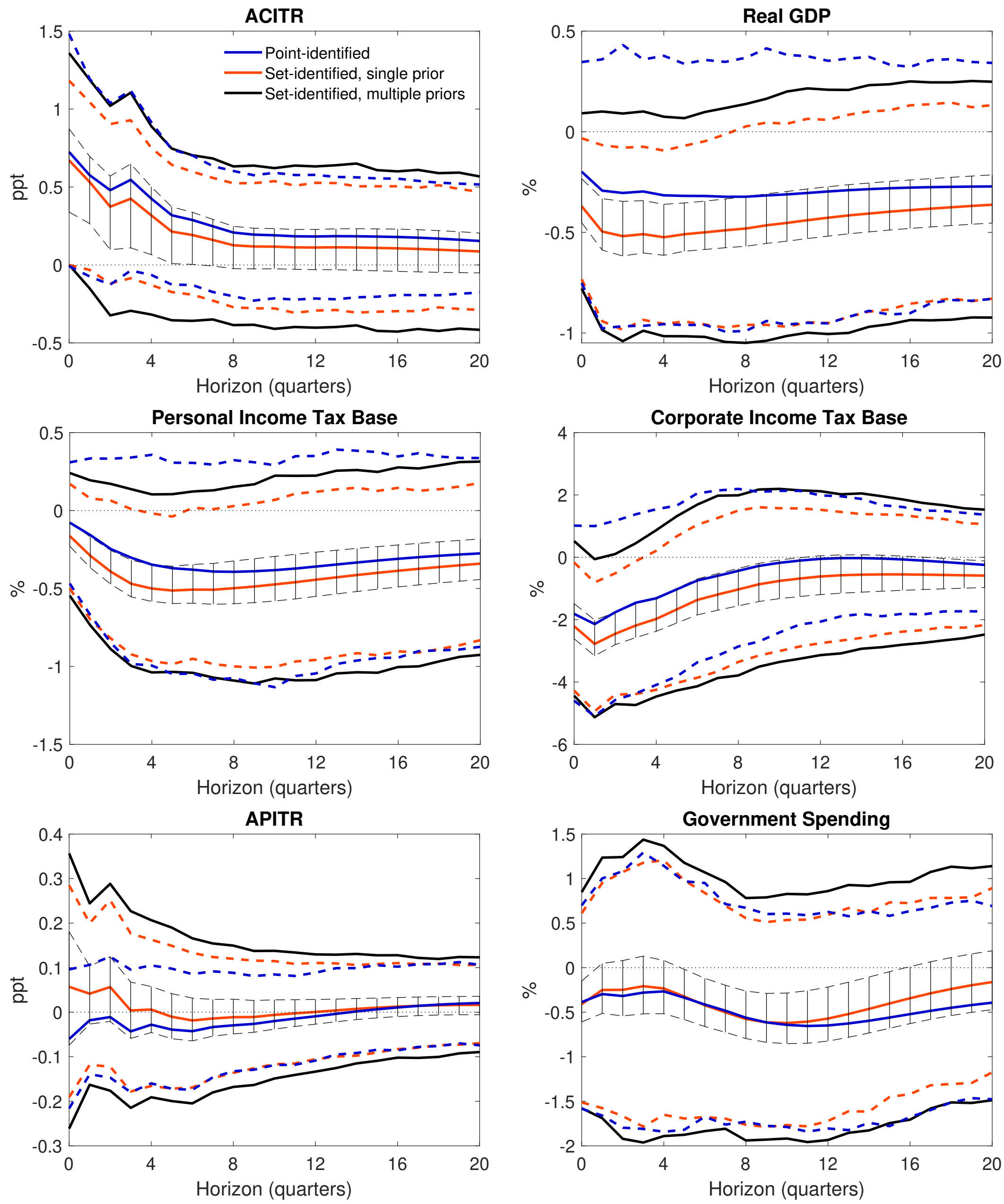
[22] The impulse responses of government debt are omitted for brevity. Note that the posterior mean and credible intervals under the point-identifying restrictions do not necessarily lie within the set of posterior means and robust credible intervals, respectively. One reason for this is because parameter values satisfying the point-identifying restrictions do not necessarily satisfy the set-identifying restrictions; for example, under the point-identifying restrictions used to identify $\varepsilon_{ACITR,t}$, the posterior probability that our sign restrictions are satisfied is around 45 per cent. Another reason is that the robust credible interval is not a union of the highest posterior density intervals over the class of posteriors.

Figure 5.1: Impulse Responses to APITR Shock



Notes: Coloured solid lines are posterior means and coloured dashed lines are associated 90 per cent highest posterior density credible intervals; vertical bars represent the set of posterior means and black solid lines are 90 per cent robust credible intervals.

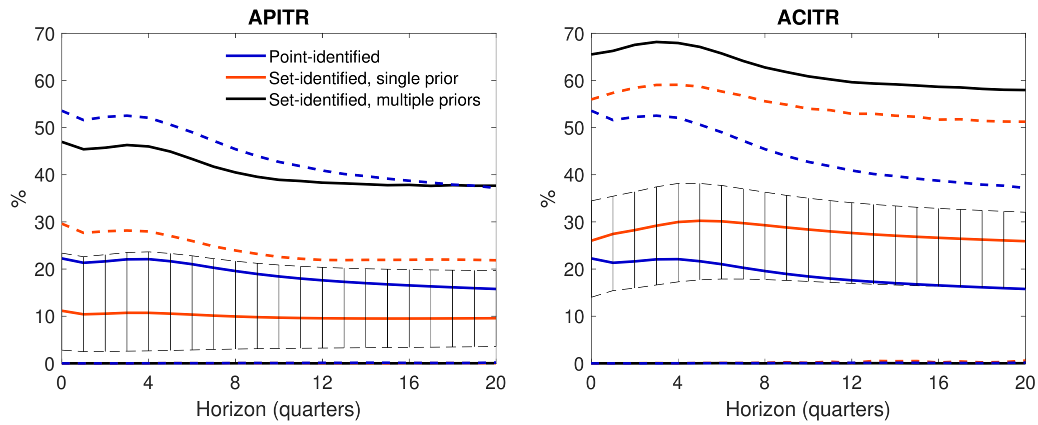
Figure 5.2: Impulse Responses to ACITR Shock



Notes: Coloured solid lines are posterior means and coloured dashed lines are associated 90 per cent highest posterior density credible intervals; vertical bars represent the set of posterior means and black solid lines are 90 per cent robust credible intervals.

around 20 per cent of the forecast error variance of output at the one-year horizon, which is similar to the contribution of the APITR under the same identifying restrictions. This contribution rises to 26 per cent under the set-identifying restrictions and the single prior. The set of posterior means ranges from 15 to 35 per cent, which suggests that ACITR shocks explain a nontrivial share of the unexpected variation in output at short horizons regardless of the choice of prior.

Figure 5.3: Contribution of Tax Shocks to Forecast Error Variance of Real GDP



Notes: Coloured solid lines are posterior means and coloured dashed lines are associated 90 per cent highest posterior density credible intervals; vertical bars represent the set of posterior means and black solid lines are 90 per cent robust credible intervals.

Since our set-identifying restrictions include both zero and sign restrictions, the identified set may be empty at particular draws of $\boldsymbol{\phi}$. The posterior probability that the identified set is non-empty, $\pi_{\boldsymbol{\phi}|\mathbf{Y},\mathbf{M}}(\{\boldsymbol{\phi} : IS_{\eta_{i,j,h}}(\boldsymbol{\phi}|F,S) \neq \emptyset\})$, is over 90 per cent, which suggests that the identifying restrictions are consistent with the data.

5.6 Conclusion

This paper develops algorithms for robust Bayesian inference in proxy SVARs where the impulse responses or FEVDs of interest are set-identified. This approach allows researchers to relax potentially controversial point-identifying restrictions without having to specify a single, unrevisable prior over the model's set-identified parameters. This is likely to be of particular value in proxy SVARs where more than one proxy is used to identify more than one structural shock.

Chapter 6

Identification and Inference Under Narrative Restrictions

6.1 Introduction

Estimating the dynamic causal effects of structural shocks is a key challenge in macroeconomics. A common approach to this problem is to use a structural vector autoregression (SVAR) with sign or zero restrictions on the model's structural parameters. Recently, a number of papers have augmented these restrictions with restrictions that involve the values of the structural shocks in specific periods. For example, Antolín-Díaz and Rubio-Ramírez (2018) (AR18) propose restricting the signs of structural shocks and their contributions to the change in particular variables in certain historical episodes. Ludvigson, Ma and Ng (2018) independently propose restricting the sign or magnitude of the structural shocks in specific periods. A burgeoning empirical literature has adopted similar restrictions, including Ben Zeev (2018), Furlanetto and Robstad (2019), Cheng and Yang (2020), Inoue and Kilian (2020), Kilian and Zhou (2020a, 2020b), Laumer (2020), Redl (2020), Zhou (2020) and Ludvigson, Ma and Ng (2020). The fact that these restrictions are placed on the shocks rather than the parameters raises novel problems related to identification, estimation and inference. This paper clarifies the nature of these problems and proposes a solution that is valid from both Bayesian and frequentist perspectives.

Henceforth, we refer to any restrictions that can be written as inequalities involving structural shocks in particular periods as 'narrative restrictions' (NR). An example of NR are 'shock-sign restrictions', such as the restriction in AR18 that the US economy was hit by a positive monetary policy shock in October 1979. This is when the Federal Reserve markedly increased the federal funds rate following Paul Volcker becoming chairman, and is widely considered an example of a positive monetary policy shock (e.g., Romer and Romer (1989)). AR18 also consider 'historical-decomposition restrictions', such as the restriction that the change in the federal funds rate in October 1979 was overwhelmingly due to a monetary policy shock. This is an inequality restriction that simultaneously constrains the historical decomposition of the federal funds rate with respect to all structural shocks in the SVAR. Other restrictions on the structural shocks also fit into this framework. For example, we additionally consider 'shock-rank restrictions', such as the restriction that the monetary policy shock in October 1979 was the largest positive realization of this shock in the sample period.

From a frequentist perspective, NR are fundamentally different from traditional identifying restrictions, such as sign restrictions on impulse responses (e.g., Uhlig (2005)). Under normally distributed structural shocks, traditional sign restrictions induce set-identification, because they generate a set-valued mapping from the SVAR's reduced-form parameters to its structural parameters that represents observational equivalence (i.e., an identified set). This set-valued mapping corresponds to the flat region of the structural-parameter likelihood and, by the definition of observational equivalence (e.g., Rothenberg (1971)), does not depend on the realization of the data. NR also result in the structural-parameter likelihood possessing flat regions and hence generate a set-valued mapping from the reduced-form parameters to the structural parameters. Crucially, this mapping depends not only on the reduced-form parameters, but also on the realization of the data. The data-dependence of this mapping implies that the standard concept of an identified set does not apply. In turn, this means that: 1) it is unclear whether NR are point- or set-identifying restrictions; and 2) there is no known valid frequentist procedure to conduct inference in these models.[1]

From a Bayesian perspective, AR18 and the empirical papers that adopt their approach conduct standard (single-prior) Bayesian inference under NR in much the same way as under traditional sign restrictions. However, we highlight two features of this approach that can spuriously affect inference. First, the *conditional* likelihood used by AR18 to construct the posterior (distribution) implies that, for some types of NR, a component of the prior (distribution) is updated only in the direction that makes the NR unlikely to hold ex ante. This occurs because the numerator of the conditional likelihood – the likelihood of the reduced-form VAR – is flat with respect to the orthonormal matrix that maps reduced-form VAR innovations into structural shocks, whereas the denominator – the ex ante probability that the NR hold – depends on this matrix. Second, standard Bayesian inference under NR may be sensitive to the choice of prior when the NR yield a likelihood with flat regions. A flat likelihood implies that the conditional posterior of the orthonormal matrix is proportional to its conditional prior whenever the likelihood is nonzero. Posterior inference may therefore be sensitive to the choice of conditional prior for the orthonormal matrix. This is a problem that also occurs in set-identified models under traditional restrictions (e.g., Poirier (1998)).

To address the above issues, we study identification under NR and propose a framework for conducting estimation and inference that is potentially appealing to both Bayesians and frequentists. We proceed in four main steps. First, we formalize the identification problem under NR. Second, we propose a simple modification of the existing Bayesian approach that eliminates the source of posterior distortion arising under NR. The modification is to use the *unconditional* likelihood, rather than the conditional likelihood, to construct the posterior. Third, as a tool for assessing and/or eliminating posterior sensitivity occurring due to the likelihood having flat regions, we propose a robust (multiple-prior) Bayesian approach to estimation and inference. Finally, we show that the robust Bayesian approach has frequentist validity in large samples.

[1] Ludvigson et al. (2018, 2021) conduct inference using a bootstrap procedure, but its frequentist validity is unknown.

To the best of our knowledge, this is the first paper to formally study identification under general NR. Plagborg-Møller and Wolf (2021b) suggest that shock-sign restrictions could in principle be recast as an external instrument (or 'proxy') and used to point-identify impulse responses in a proxy SVAR or local projection framework. However, it is unclear how richer sets of NR, including restrictions on the historical decomposition, could be cast as proxies without discarding identifying information. Petterson, Seim and Shapiro (2020) derive bounds for a slope parameter in a single equation given restrictions on the plausible magnitude of the residuals, but the restrictions are over the entire sample and the setting is non-probabilistic.

We make two main contributions to the study of identification under NR. First, we provide a necessary and sufficient condition for global identification of an SVAR under NR and show that this condition is satisfied in a simple bivariate example with a single shock-sign restriction. That is, in contrast with traditional sign restrictions, NR may be formally point-identifying despite generating a set-valued mapping from reduced-form to structural parameters in any particular sample. However, this point-identification result does not deliver a point estimator, because the observed likelihood is almost always flat at the maximum. Second, to develop a frequentist-valid procedure for inference, we introduce the notion of a 'conditional identified set'. The conditional identified set extends the standard notion of an identified set to a setting where identification is defined in a repeated sampling experiment conditional on the set of observations entering the NR. This provides an interpretation for the set-valued mapping induced by the NR as the set of observationally equivalent structural parameters in such a conditional frequentist experiment.

The feature of having a set of maximum likelihood estimators is similar to maximum score estimation, where the maximum score objective function yields a set of maximizers (Manski (1975, 1985)). Our conditional identified set, which fixes the flat regions of the likelihood in the conditional frequentist experiment, is analogous to the finite-sample identified set introduced by Rosen and Ura (2020) in the maximum score context; however, their finite-sample inference approach does not apply to the current setting.

In terms of inference under NR, this paper makes contributions from both a Bayesian and a frequentist point of view.

The paper's contribution to Bayesian inference is to address the issues associated with the current approach to standard Bayesian inference under NR. First, we advocate using the *unconditional* likelihood – the joint probability of observing the data and the NR being satisfied – when constructing the posterior, rather than the conditional likelihood. Regardless of the type of NR imposed, the unconditional likelihood is flat with respect to the orthonormal matrix that maps reduced-form VAR innovations into structural shocks. This removes the source of posterior distortion that arises due to conditioning on the NR holding. Standard Bayesian inference under the unconditional likelihood requires a simple change to existing computational algorithms. Second, to address posterior sensitivity to the choice of prior, we adapt the robust Bayesian approach of Giacomini and Kitagawa (2021) (GK) to a setting with NR.

In the context of an SVAR under traditional identifying restrictions, the robust Bayesian approach of GK involves decomposing the prior for the structural parameters into a prior for the reduced-form parameters, which is revised by the data, and a conditional prior for the orthonormal matrix given the reduced-form parameters, which is unrevisable. Considering the class of all conditional priors for the orthonormal matrix that are consistent with the identifying restrictions generates a class of posteriors, which can be summarized by a set of posterior means (an estimator of the identified set) and a robust credible region. This removes the source of posterior sensitivity.[2]

We show that this approach can also be used to summarize posterior sensitivity under NR, since the unconditional likelihood at the realized data possesses flat regions and the posterior can therefore be sensitive to the choice of prior, as in standard set-identified models. There are, however, some modifications needed to account for the novel features of the NR. In particular, one cannot use a conditional prior for the orthonormal matrix to impose the NR due to the data-dependent map-

[2]Giacomini, Kitagawa and Read (2022) extend this approach to proxy SVARs where the parameters of interest are set-identified using external instruments.

ping between reduced-form and structural parameters. However, by considering the class of all conditional priors consistent with any traditional identifying restrictions (if present), one can trace out all possible posteriors that are consistent with the traditional restrictions *and* the NR. This is because traditional restrictions truncate the support of the conditional prior, while NR truncate the support of the likelihood. Consequently, the posterior given any particular conditional prior is only supported on the common support of the conditional prior and the likelihood.

If the researcher has a credible conditional prior, we recommend reporting the standard Bayesian posterior under the unconditional likelihood together with the robust Bayesian output. This allows other researchers to assess the extent to which posterior inference may be driven by prior choice. In the absence of a credible conditional prior, the robust Bayesian output should be reported as an alternative to the standard Bayesian posterior.

The paper's contribution to frequentist inference is to provide an asymptotically valid approach to inference under NR, which, to the best of our knowledge, was not previously available. To explore the asymptotic frequentist properties of our robust Bayesian procedure, we assume a fixed number of NR. This assumption is empirically relevant given that applications typically impose no more than a handful of NR. We provide conditions under which the robust credible region provides asymptotically valid frequentist coverage of the conditional identified set for the impulse response. Since the conditional identified set is guaranteed to include the true impulse response, the robust credible region also provides valid coverage of the true impulse response. Our robust Bayesian approach should therefore appeal to Bayesians as well as frequentists.

We illustrate our methods by estimating the effects of monetary policy shocks in the United States. We find that posterior inferences about the response of output obtained under restrictions based on the October 1979 episode may be sensitive to the choice of conditional prior for the orthonormal matrix. In contrast, under an extended set of restrictions constructed by AR18 based on multiple historical episodes, output falls with high posterior probability following a positive monetary policy shock regardless of the choice of conditional prior. We also estimate the set of output responses that are consistent with the restriction that the monetary policy shock in October 1979 was the largest positive realization of the shock in the sample period. Compared with the extended set of restrictions, this shock-rank restriction results in broadly similar robust posterior inferences about the output response.

Outline. The remainder of the paper is structured as follows. Section 6.2 highlights the econometric issues that arise when imposing NR using a simple bivariate example. Section 6.3 describes the general SVAR(p) framework. Section 6.4 formally analyzes identification under NR and introduces the concept of a conditional identified set. Section 6.5 discusses how to conduct standard and robust Bayesian inference under NR. Section 6.6 explores the frequentist properties of the robust Bayesian approach. Section 6.7 contains the empirical application and Section 6.8 concludes. The appendices contain proofs and other supplemental material.

Generic notation: For the matrix $\mathbf{X}$, $\text{vec}(\mathbf{X})$ is the vectorization of $\mathbf{X}$ and $\text{vech}(\mathbf{X})$ is the half-vectorization of $\mathbf{X}$ (when $\mathbf{X}$ is symmetric). $\mathbf{e}_{i,n}$ is the ith column of the $n \times n$ identity matrix, $\mathbf{I}_n$. $\mathbf{0}_{n\times m}$ is a $n \times m$ matrix of zeros. $1(.)$ is the indicator function. $\|.\|$ is the Euclidean norm.

6.2 Bivariate Example

This section sets out the econometric issues that arise when imposing NR using the simplest possible SVAR as an example. Consider the SVAR(0) $\mathbf{A}_0\mathbf{y}_t = \boldsymbol{\varepsilon}_t$, for $t = 1,\ldots,T$, where $\mathbf{y}_t = (y_{1t}, y_{2t})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ with $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$. We abstract from dynamics for ease of exposition, but this is without loss of generality. The orthogonal reduced form of the model reparameterizes $\mathbf{A}_0$ as $\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$, where $\boldsymbol{\Sigma}_{tr}$ is the lower-triangular Cholesky factor (with positive diagonal elements) of $\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{y}_t\mathbf{y}_t') = \mathbf{A}_0^{-1}\left(\mathbf{A}_0^{-1}\right)'$. We parameterize $\boldsymbol{\Sigma}_{tr}$ directly as

$$\boldsymbol{\Sigma}_{tr} = \begin{bmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \quad (\sigma_{11}, \sigma_{22} > 0), \tag{6.1}$$

and denote the vector of reduced-form parameters as $\boldsymbol{\phi} = \text{vech}(\boldsymbol{\Sigma}_{tr})$. $\mathbf{Q}$ is an orthonormal matrix in the space of 2×2 orthonormal matrices, $\mathscr{O}(2)$:

$$\mathbf{Q} \in \mathscr{O}(2) = \left\{ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\} \cup \left\{ \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} : \theta \in [-\pi, \pi] \right\}, \quad (6.2)$$

where the first set is the set of 'rotation' matrices and the second set is the set of 'reflection' matrices. Henceforth, we leave the restriction $\theta \in [-\pi, \pi]$ implicit.

The set of values for $\mathbf{A}_0$ that are consistent with the reduced-form parameters in the absence of additional restrictions is

$$\mathbf{A}_0 \in \left\{ \frac{1}{\sigma_{11}\sigma_{22}} \begin{bmatrix} \sigma_{22}\cos\theta - \sigma_{21}\sin\theta & \sigma_{11}\sin\theta \\ -\sigma_{21}\cos\theta - \sigma_{22}\sin\theta & \sigma_{11}\cos\theta \end{bmatrix} \right\} \cup \left\{ \frac{1}{\sigma_{11}\sigma_{22}} \begin{bmatrix} \sigma_{22}\cos\theta - \sigma_{21}\sin\theta & \sigma_{11}\sin\theta \\ \sigma_{22}\sin\theta + \sigma_{21}\cos\theta & -\sigma_{11}\cos\theta \end{bmatrix} \right\}. \quad (6.3)$$

Below, we always impose the 'sign normalization'that $\text{diag}(\mathbf{A}_0) \geq \mathbf{0}_{2\times 1}$, which is a normalization on the signs of the structural shocks.

6.2.1 Shock-sign restrictions

Consider the 'shock-sign restriction' that ε_{1k} is nonnegative for some $k \in \{1, \ldots, T\}$:

$$\varepsilon_{1k} = \mathbf{e}_{1,2}'\mathbf{A}_0\mathbf{y}_k = (\sigma_{11}\sigma_{22})^{-1}(\sigma_{22}y_{1k}\cos\theta + (\sigma_{11}y_{2k} - \sigma_{21}y_{1k})\sin\theta) \geq 0. \quad (6.4)$$

Given the realization of the data in period k, Equation (6.4) implies that the restricted structural shock can be written as a function $\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k)$. Under the sign normalization and the shock-sign restriction, θ is restricted to the set

$$\theta \in \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \geq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\}$$
$$\cup \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \leq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\}. \tag{6.5}$$

Since y_{1k} and y_{2k} enter the inequalities characterising this set, the shock-sign restriction induces a set-valued mapping from $\boldsymbol{\phi}$ to θ that depends on the realization of $\mathbf{y}_k$. For example, if $\sigma_{21} < 0$, $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$ and $y_{1k} > 0$,

$$\theta \in \left[\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right)\right].[3] \tag{6.6}$$

The direct dependence of this mapping on the realization of the data implies that the standard notion of an identified set – the set of observationally equivalent structural parameter values given the reduced-form parameters – does not apply. Consequently, it is not obvious whether existing frequentist procedures for conducting inference in set-identified models are valid under NR. Moreover, it is unclear whether the restrictions are, in fact, set-identifying in a formal frequentist sense. We formally analyze identification under NR in Section 6.4.

When conducting Bayesian inference, AR18 construct the posterior using the conditional likelihood, which is the likelihood of observing the data conditional on the NR holding. Letting $\mathbf{y}^T = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$ represent a realization of the random variable $\mathbf{Y}^T$, the conditional likelihood is

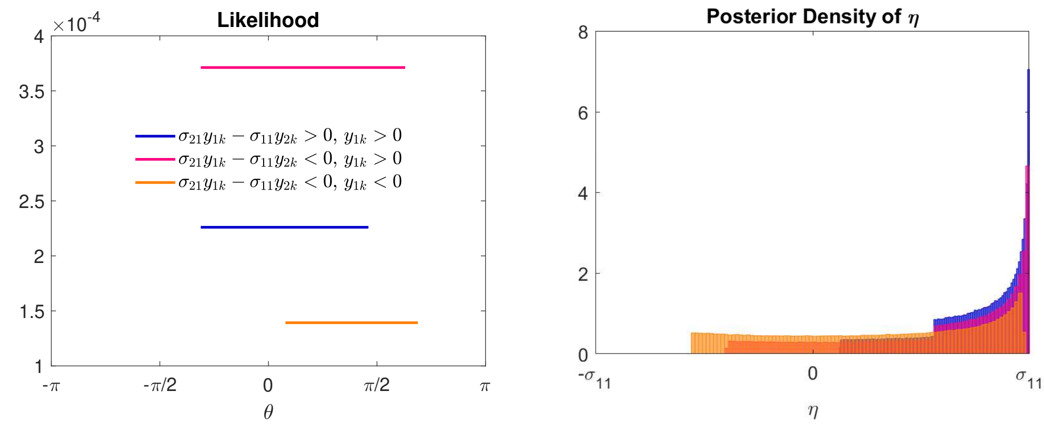
$$p\left(\mathbf{y}^T|\theta, \boldsymbol{\phi}, \varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0\right) = \frac{\prod_{t=1}^T (2\pi)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\mathbf{y}_t'\boldsymbol{\Sigma}^{-1}\mathbf{y}_t\right)}{\Pr(\varepsilon_{1k} \geq 0|\theta, \boldsymbol{\phi})} 1\left(\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0\right). \tag{6.7}$$

The numerator in the first term is a function of $\boldsymbol{\phi}$ and $\mathbf{y}^T$, while the denominator is equal to 1/2, because the marginal distribution of ε_{1k} is standard normal. The conditional likelihood therefore depends on θ only through the indicator function

[3] See Appendix D.1 for the full characterization of this mapping.

$1(\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0)$. This indicator function truncates the likelihood, with the truncation points depending on $\mathbf{y}_k$. To illustrate, the left panel of Figure 6.1 plots the likelihood given different realizations of the data drawn from a data-generating process with $\sigma_{21} < 0$ and assuming for simplicity that the econometrician knows $\boldsymbol{\phi}$.[4] The conditional likelihood is flat over the region for θ satisfying the shock-sign restriction and is zero outside this region. The support of the nonzero region depends on the realization of $\mathbf{y}_k$.

Figure 6.1: Shock-sign Restriction



Notes: $T = 3$, $\boldsymbol{\phi}$ is known and $\varepsilon_{1k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) \geq 0$ is the narrative sign restriction; likelihood in top-left panel is zero outside of plotted intervals; posterior density of $\eta = \sigma_{11}\cos\theta$ is approximated using 1,000,000 draws of θ from its uniform posterior.

The flat likelihood function implies that the posterior will be proportional to the prior in the region where the likelihood function is nonzero, and it will be zero outside this region. The standard approach to Bayesian inference in SVARs identified via sign restrictions assumes a uniform (or Haar) prior over $\mathbf{Q}$, as does the approach in AR18.[5] In the bivariate example, this is equivalent to a prior for θ that is uniform over the interval $[-\pi, \pi]$. This prior implies that the posterior for θ is also uniform over the interval for θ where the likelihood function is nonzero.

[4]The data-generating process assumes $\mathbf{A}_0 = \begin{bmatrix} 1 & 0.5 \\ 0.2 & 1.2 \end{bmatrix}$, which implies that $\theta = \arcsin(0.5\sigma_{22})$ with $\mathbf{Q}$ equal to the rotation matrix. We assume the time series is of length $T = 3$ and draw sequences of structural shocks such that $\varepsilon_{1,1} \geq 0$. T is a small number to control Monte Carlo sampling error in the exercises below. The analysis with known $\boldsymbol{\phi}$ replicates the situation with a large sample, where the likelihood for $\boldsymbol{\phi}$ concentrates at the truth. The assumption that $\boldsymbol{\phi}$ is known also facilitates visualizing the likelihood, which otherwise is a function of four parameters.

[5]See, for example, Uhlig (2005), Rubio-Ramírez, Waggoner and Zha (2010), Baumeister and Hamilton (2015) and Arias, Rubio-Ramírez and Waggoner (2018).

The impact impulse response of y_{1t} to a positive standard-deviation shock ε_{1t} is $\eta \equiv \sigma_{11}\cos\theta$. The right panel of Figure 6.1 plots the posterior for η induced by a uniform prior over θ given the same realizations of the data for which the likelihood was plotted in the left panel. The uniform posterior for θ induces a posterior for η that assigns more probability mass to more-extreme values of η. This highlights that even a 'uniform' prior may be informative for parameters of interest, which is also the case under traditional sign restrictions (Baumeister and Hamilton (2015)). One difference is that the conditional prior under sign restrictions is never updated by the data, whereas the support and shape of the posterior for η under NR may depend on the realization of $\mathbf{y}_k$ through its effect on the truncation points of the likelihood, so there may be some updating of the conditional prior by the data. For example, when $\sigma_{21} < 0$, $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$ and $y_{1k} > 0$,

$$\eta \in \left[\sigma_{11}\cos\left(\arctan\left(\max\left\{-\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right)\right), \sigma_{11}\right]. \tag{6.8}$$

However, the conditional prior is not updated at values of θ corresponding to the flat region of the likelihood. Posterior inference about η may therefore still be sensitive to the choice of prior, as in standard set-identified SVARs.

6.2.2 Historical-decomposition restrictions

The historical decomposition is the contribution of a particular structural shock to the observed unexpected change in a particular variable over some horizon. The contribution of the first shock to the change in the first variable in the kth period is

$$H_{1,1,k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) = \sigma_{22}^{-1}\left(\sigma_{22}y_{1k}\cos^2\theta + (\sigma_{11}y_{2k} - \sigma_{21}y_{1k})\cos\theta\sin\theta\right), \tag{6.9}$$

while the contribution of the second shock is

$$H_{1,2,k}(\theta, \boldsymbol{\phi}, \mathbf{y}_k) = \sigma_{22}^{-1}\left(\sigma_{22}y_{1k}\sin^2\theta + (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\cos\theta\sin\theta\right). \tag{6.10}$$

Consider the restriction that the first structural shock in period k was positive and (in the language of AR18) the 'most important contributor' to the change in the first variable, which requires that $|H_{1,1,k}(\theta,\boldsymbol{\phi},\mathbf{y}_k)| \geq |H_{1,2,k}(\theta,\boldsymbol{\phi},\mathbf{y}_k)|$. Under these restrictions and the sign normalization, θ must satisfy a set of inequalities that depends on $\boldsymbol{\phi}$ and $\mathbf{y}_k$. As in the case of the shock-sign restriction, this set of restrictions generates a set-valued mapping from $\boldsymbol{\phi}$ to θ that depends on $\mathbf{y}_k$.[6]

Let $\mathscr{D}(\theta,\boldsymbol{\phi},\mathbf{y}_k) = 1\{\varepsilon_{1k}(\theta,\boldsymbol{\phi},\mathbf{y}_k) \geq 0, |H_{1,1,k}(\theta,\boldsymbol{\phi},\mathbf{y}_k)| \geq |H_{1,2,k}(\theta,\boldsymbol{\phi},\mathbf{y}_k)|\}$ represent the indicator function equal to one when the NR are satisfied and equal to zero otherwise, and let $\tilde{\mathscr{D}}(\theta,\boldsymbol{\phi},\boldsymbol{\varepsilon}_k) = 1\{\varepsilon_{1k} \geq 0, |\tilde{H}_{1,1,k}(\theta,\boldsymbol{\phi},\varepsilon_{1k})| \geq |\tilde{H}_{1,2,k}(\theta,\boldsymbol{\phi},\varepsilon_{2k})|\}$ represent the indicator function for the same event in terms of the structural shocks rather than the data. The conditional likelihood function given the restrictions is then

$$p\left(\mathbf{y}^T|\theta,\boldsymbol{\phi},\mathscr{D}(\theta,\boldsymbol{\phi},\mathbf{y}_k)=1\right) = \frac{\prod_{t=1}^T (2\pi)^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\mathbf{y}_t'\boldsymbol{\Sigma}^{-1}\mathbf{y}_t\right)}{\Pr(\tilde{\mathscr{D}}(\theta,\boldsymbol{\phi},\boldsymbol{\varepsilon}_k)=1|\theta,\boldsymbol{\phi})}\mathscr{D}(\theta,\boldsymbol{\phi},\mathbf{y}_k). \tag{6.11}$$

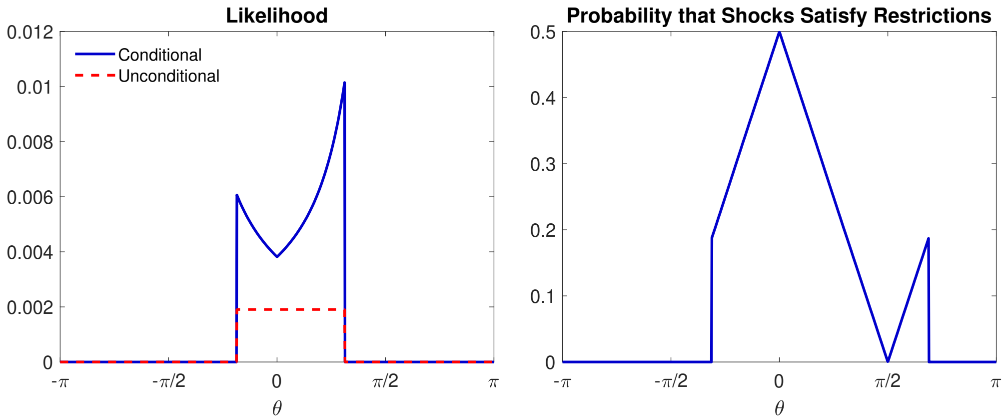
As in the case of the shock-sign restriction, the numerator of the first term does not depend on θ. In contrast, the probability in the denominator now depends on θ through the historical decomposition. Intuitively, changing θ changes the impulse responses of y_{1t} to the two shocks and thus changes the ex ante probability that $|\tilde{H}_{1,1,k}(\theta,\boldsymbol{\phi},\varepsilon_{1k})| \geq |\tilde{H}_{1,2,k}(\theta,\boldsymbol{\phi},\varepsilon_{2k})|$. The conditional likelihood therefore depends on θ both through this probability and through the indicator function determining the truncation points of the likelihood. Consequently, the likelihood function is not necessarily flat when it is nonzero.

To illustrate, the left panel of Figure 6.2 plots the conditional likelihood evaluated at a random realization of the data satisfying the restrictions using the same data-generating process as above and assuming that $\boldsymbol{\phi}$ is known. The probability in the denominator of the conditional likelihood is approximated by drawing 1,000,000 realizations of $\boldsymbol{\varepsilon}_k$ and computing the proportion of draws satisfying the restrictions at each value of θ. This probability is plotted in the right panel of Figure 6.2. The likelihood is again truncated according to a set-valued mapping from $\boldsymbol{\phi}$ and $\mathbf{y}_k$ to θ, but an important difference from the case with the shock-sign re-

[6] It is more difficult to analytically characterize the induced mapping than in the shock-sign example, so we do not pursue this.

striction is that the likelihood is no longer flat within the region where it is nonzero. In particular, the conditional likelihood has a maximum at the value of θ that minimizes the ex ante probability that the NR are satisfied (within the set of values of θ that are consistent with the restrictions). The posterior for θ induced by a uniform prior will therefore assign greater posterior probability to values of θ that yield a lower ex ante probability of satisfying the NR.

Figure 6.2: Historical-decomposition Restriction



Notes: $T = 3$ and $\boldsymbol{\phi}$ is known; $\varepsilon_{1,1}(\boldsymbol{\phi}, \theta, \mathbf{y}_k) \geq 0$ and $|H_{1,1,1}(\boldsymbol{\phi}, \theta, \mathbf{y}_k)| \geq |H_{2,1,1}(\boldsymbol{\phi}, \theta, \mathbf{y}_k)|$ are the narrative sign restrictions; $\Pr(\tilde{\mathscr{D}}(\theta, \boldsymbol{\phi}, \boldsymbol{\varepsilon}_k) = 1|\theta, \boldsymbol{\phi})$ is approximated using 1,000,000 Monte Carlo draws.

If we view the narrative event as a part of the observables and its probability of occurring depends on the parameter of interest, conditioning on the narrative event implies that we are conditioning on a non-ancillary statistic. When conducting likelihood-based inference, conditioning on a non-ancillary statistic is undesirable, because it represents a loss of information about the parameter of interest. The probability that the shock-sign restriction is satisfied is independent of the parameters, so the event that the restriction is satisfied is ancillary. In the case where there is also a restriction on the historical decomposition, the probability that the NR are satisfied depends on θ, so the event that the NR are satisfied is not ancillary. Conditioning on this non-ancillary event results in the likelihood no longer being flat, but the shape of the likelihood is fully driven by the inverse probability of the conditioning event. That is, the loss of information for θ can be viewed as distorting the shape of the posterior in the sense that the prior is updated toward values of θ that make the event that the NR are satisfied less likely ex ante. We therefore advocate forming the likelihood without conditioning on the restrictions holding.

The joint (or unconditional) likelihood of observing the data and the NR holding is obtained by multiplying the conditional likelihood by the probability that the NR are satisfied:

$$p\left(\mathbf{y}^T, \tilde{\mathscr{D}}(\theta, \boldsymbol{\phi}, \boldsymbol{\varepsilon}_k) = 1 | \theta, \boldsymbol{\phi}\right) = \prod_{t=1}^{T} (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t\right)\right) \mathscr{D}(\theta, \boldsymbol{\phi}, \mathbf{y}_k). \tag{6.12}$$

Conditional on being nonzero, the unconditional likelihood is flat with respect to θ. The unconditional likelihood depends on θ only through the points of truncation. To illustrate, Figure 6.2 plots the unconditional likelihood given the same realization of the data used to plot the conditional likelihood. As in the case of the shock-sign restriction, the flat unconditional likelihood implies that posterior inference may be sensitive to the choice of prior. We describe our approach to addressing this posterior sensitivity in Section 6.5.2.

6.3 General Framework

This section describes the general SVAR(p) and outlines the restrictions that we consider.

6.3.1 SVAR(p)

Let $\mathbf{y}_t$ be an $n \times 1$ vector of endogenous variables following the SVAR(p) process:

$$\mathbf{A}_0 \mathbf{y}_t = \sum_{l=1}^{p} \mathbf{A}_l \mathbf{y}_{t-l} + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{6.13}$$

where $\mathbf{A}_0$ is invertible and $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{n\times 1}, \mathbf{I}_n)$ are structural shocks. The initial conditions $(\mathbf{y}_{1-p}, ..., \mathbf{y}_0)$ are given. We omit exogenous regressors (such as a constant) for simplicity of exposition, but these are straightforward to include. Letting $\mathbf{x}_t = (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')'$ and $\mathbf{A}_+ = (\mathbf{A}_1, \ldots, \mathbf{A}_p)$, rewrite the SVAR($p$) as

$$\mathbf{A}_0 \mathbf{y}_t = \mathbf{A}_+ \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T. \tag{6.14}$$

$(\mathbf{A}_0, \mathbf{A}_+)$ are the structural parameters. The reduced-form VAR(p) representation is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u}_t, \quad t = 1, ..., T, \tag{6.15}$$

where $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_p)$, $\mathbf{B}_l = \mathbf{A}_0^{-1}\mathbf{A}_l$ for $l = 1, \ldots, p$, and $\mathbf{u}_t = \mathbf{A}_0^{-1}\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{n\times 1}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$. $\boldsymbol{\phi} = (\mathrm{vec}(\mathbf{B})', \mathrm{vech}(\boldsymbol{\Sigma})')' \in \boldsymbol{\Phi}$ are the reduced-form parameters. We assume that $\mathbf{B}$ is such that the VAR(p) can be inverted into an infinite-order vector moving average (VMA(∞)) representation.[7]

As is standard in the literature that considers set-identified SVARs, we reparameterize the model into its orthogonal reduced form (e.g., Arias et al. (2018)):

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \boldsymbol{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{6.16}$$

where $\boldsymbol{\Sigma}_{tr}$ is the lower-triangular Cholesky factor of $\boldsymbol{\Sigma}$ (i.e. $\boldsymbol{\Sigma}_{tr}\boldsymbol{\Sigma}_{tr}' = \boldsymbol{\Sigma}$) with diagonal elements normalized to be non-negative, $\mathbf{Q}$ is an $n \times n$ orthonormal matrix and $\mathscr{O}(n)$ is the set of all such matrices. The structural and orthogonal reduced-form parameterizations are related through the mapping $\mathbf{B} = \mathbf{A}_0^{-1}\mathbf{A}_+$, $\boldsymbol{\Sigma} = \mathbf{A}_0^{-1}(\mathbf{A}_0^{-1})'$ and $\mathbf{Q} = \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{A}_0^{-1}$ with inverse mapping $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$ and $\mathbf{A}_+ = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{B}$.

The VMA(∞) representation of the model is

$$\mathbf{y}_t = \sum_{h=0}^{\infty} \mathbf{C}_h \mathbf{u}_{t-h} = \sum_{h=0}^{\infty} \mathbf{C}_h \boldsymbol{\Sigma}_{tr}\mathbf{Q}\boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{6.17}$$

where $\mathbf{C}_h$ is the hth term in $(\mathbf{I}_n - \sum_{l=1}^{p}\mathbf{B}_l L^l)^{-1}$ and L is the lag operator. $\mathbf{C}_h$ is defined recursively by $\mathbf{C}_h = \sum_{l=1}^{\min\{k,p\}}\mathbf{B}_l\mathbf{C}_{h-l}$ for $h \geq 1$ with $\mathbf{C}_0 = \mathbf{I}_n$. The (i, j)th element of the matrix $\mathbf{C}_h\boldsymbol{\Sigma}_{tr}\mathbf{Q}$, which we denote by $\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q})$, is the horizon-$h$ impulse response of the ith variable to the jth structural shock:

$$\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{e}_{i,n}'\mathbf{C}_h\boldsymbol{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_{j,n} = \mathbf{c}_{i,h}'(\boldsymbol{\phi})\mathbf{q}_j, \tag{6.18}$$

where $\mathbf{c}_{i,h}'(\boldsymbol{\phi}) = \mathbf{e}_{i,n}'\mathbf{C}_h\boldsymbol{\Sigma}_{tr}$ is the ith row of $\mathbf{C}_h\boldsymbol{\Sigma}_{tr}$ and $\mathbf{q}_j = \mathbf{Q}\mathbf{e}_{j,n}$ is the jth column

[7] The VAR(p) is invertible into a VMA(∞) process when the eigenvalues of the companion matrix lie inside the unit circle. See Hamilton (1994) or Kilian and Lütkepohl (2017).

of $\mathbf{Q}$.

6.3.2 Narrative restrictions

In the absence of any identifying restrictions, it is well-known that $\mathbf{Q}$ is set-identified. Consequently, functions of $\mathbf{Q}$, such as the impulse responses, are also set-identified. Imposing traditional identifying restrictions on the SVAR is equivalent to restricting $\mathbf{Q}$ to lie in a subspace of $\mathscr{O}(n)$. It is conventional to impose a 'sign normalization' on the structural shocks. We normalize the diagonal elements of $\mathbf{A}_0$ to be non-negative, so a positive value of ε_{it} is a positive shock to the ith equation in the SVAR at time t. The sign normalization implies that $\text{diag}(\mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times 1}$.

It is common to impose sign restrictions on the impulse responses (e.g., Uhlig (2005)) or on the structural parameters themselves. For example, the restriction that the horizon-h impulse response of the ith variable to the jth shock is nonnegative is $c'_{i,h}(\boldsymbol{\phi})\mathbf{q}_j \geq 0$, which is a linear inequality restriction on a single column of $\mathbf{Q}$ that depends only on the reduced-form parameter $\boldsymbol{\phi}$. Restrictions on elements of $\mathbf{A}_0$ take a similar form.

In contrast, NR constrain the values of the structural shocks in particular periods. The structural shocks are

$$\boldsymbol{\varepsilon}_t = \mathbf{A}_0\mathbf{u}_t = \mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}\mathbf{u}_t. \tag{6.19}$$

The shock-sign restriction that the ith structural shock at time k is positive is

$$\varepsilon_{ik}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_k) = \mathbf{e}'_{i,n}\mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}\mathbf{u}_k = (\mathbf{\Sigma}_{tr}^{-1}\mathbf{u}_k)'\mathbf{q}_i \geq 0. \tag{6.20}$$

We can treat $\mathbf{u}_t$ as observable given $\boldsymbol{\phi}$ and the data, so we suppress the dependence of $\mathbf{u}_t$ on $\boldsymbol{\phi}$ and $(\mathbf{y}'_t,\mathbf{x}'_t)'$ for notational convenience. The restriction in (6.20) is a linear inequality restriction on a single column of $\mathbf{Q}$. In contrast with traditional sign restrictions, the shock-sign restriction depends directly on the data through the reduced-form VAR innovations.

In addition to shock-sign restrictions, AR18 consider restrictions on the historical decomposition, which is the cumulative contribution of the jth shock to the observed unexpected change in the ith variable between periods k and $k+h$:

$$H_{i,j,k,k+h}\left(\boldsymbol{\phi},\mathbf{Q},\{\mathbf{u}_t\}_{t=k}^{k+h}\right) = \sum_{l=0}^{h} \mathbf{e}'_{i,n}\mathbf{C}_l\boldsymbol{\Sigma}_{tr}\mathbf{Q}\mathbf{e}_{j,n}\mathbf{e}'_{j,n}\boldsymbol{\varepsilon}_{k+h-l} \tag{6.21}$$

$$= \sum_{l=0}^{h} \mathbf{c}'_{i,l}(\boldsymbol{\phi})\mathbf{q}_j\mathbf{q}'_j\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{k+h-l}. \tag{6.22}$$

One example of a restriction on the historical decomposition is that the jth structural shock was the 'most important contributor' to the change in the ith variable between periods k and $k+h$, which requires that $|H_{i,j,k,k+h}| \geq \max_{l\neq j}|H_{i,l,k,k+h}|$. Another example is that the jth structural shock was the 'overwhelming contributor' to the change in the ith variable between periods k and $k+h$, which requires that $|H_{i,j,k,k+h}| \geq \sum_{l\neq j}|H_{i,l,k,k+h}|$. From Equation (6.21), it is clear that these restrictions are nonlinear inequality constraints that simultaneously constrain every column of $\mathbf{Q}$ and that depend on the realizations of the data in particular periods in addition to the reduced-form parameters.

Other restrictions also naturally fit into this framework. For instance, we can consider restrictions on the relative magnitudes of a particular structural shock in different periods. We refer to these restrictions as 'shock-rank restrictions', since they imply a (possibly partial) ordering of the shocks. As an example, one could impose that the ith shock in period k was the largest positive realization of this shock in the observed sample. This requires that $\varepsilon_{ik}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_k) \geq \max_{t\neq k}\{\varepsilon_{it}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_t)\}$, which can be expressed as a system of $T-1$ linear inequality restrictions on a single column of $\mathbf{Q}$: $(\boldsymbol{\Sigma}_{tr}^{-1}(\mathbf{u}_k-\mathbf{u}_t))'\mathbf{q}_i \geq 0$ for $t\neq k$. Alternatively, one could impose that the ith shock in period k was the largest-magnitude realization of that shock, or $|\varepsilon_{ik}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_k)| \geq \max_{t\neq k}\{|\varepsilon_{it}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_t)|\}$. If $\varepsilon_{ik}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_k)\geq 0$, this would require that $(\boldsymbol{\Sigma}_{tr}^{-1}(\mathbf{u}_k-\mathbf{u}_t))'\mathbf{q}_i \geq 0$ and $(\boldsymbol{\Sigma}_{tr}^{-1}(\mathbf{u}_k+\mathbf{u}_t))'\mathbf{q}_i \geq 0$ for $t\neq k$, which is a system of $2(T-1)$ linear inequalities constraining $\mathbf{q}_i$. These restrictions could also be applied to a subset of the observations rather than the full sample (e.g., $\varepsilon_{ik}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_k) > \varepsilon_{it}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_t)$ for some $t\in\{1,\ldots,T\}$).[8]

[8] Similar to the shock-rank restrictions we describe, Ben Zeev (2018) imposes a restriction on the timing of the maximum three-year average of a particular shock, as well as restrictions on the sign and relative magnitudes of this three-year average in specific periods. Restrictions on averages of

The collection of NR can be represented in the general form $N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}$, where s is the number of restrictions. As an illustration, consider the case where there is a single shock-sign restriction in period k, $\varepsilon_{1k}(\boldsymbol{\phi},\mathbf{Q},\mathbf{u}_k) \geq 0$, as well as the restriction that the first structural shock was the most important contributor to the change in the first variable in period k. Then,

$$N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) = \begin{bmatrix} (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_k)'\mathbf{q}_1 \\ |\mathbf{e}'_{1,n}\boldsymbol{\Sigma}_{tr}\mathbf{q}_1\mathbf{q}'_1\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_k| - \max_{j\neq 1}|\mathbf{e}'_{1,n}\boldsymbol{\Sigma}_{tr}\mathbf{q}_j\mathbf{q}'_j\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_k| \end{bmatrix} \geq \mathbf{0}_{2\times 1}. \tag{6.23}$$

Traditional sign and zero restrictions can also be applied alongside NR. We follow AR18 by explicitly allowing for sign restrictions on impulse responses and on elements of $\mathbf{A}_0$. We denote such sign restrictions by $S(\boldsymbol{\phi},\mathbf{Q}) \geq \mathbf{0}_{\tilde{s}\times 1}$, where $\tilde{s}$ is the number of traditional sign restrictions. It is straightforward to additionally allow for zero restrictions, including 'short-run' zero restrictions (as in Sims (1980)), 'long-run' zero restrictions (as in Blanchard and Quah (1989)), or restrictions arising from external instruments (as in Mertens and Ravn (2013) and Stock and Watson (2018)); for example, see GK and Giacomini et al. (2022).

6.3.3 Conditional and unconditional likelihoods

When constructing the posterior of the SVAR's parameters, AR18 use the likelihood conditional on the NR holding. Define

$$\begin{aligned} D_N = D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) &\equiv 1\{N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}\}, \\ r(\boldsymbol{\phi},\mathbf{Q}) &\equiv \Pr(D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) = 1|\boldsymbol{\phi},\mathbf{Q}), \\ f(\mathbf{y}^T|\boldsymbol{\phi}) &\equiv \prod_{t=1}^{T}(2\pi)^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)\right). \end{aligned}$$

The likelihood conditional on $D_N = 1$ can be written as

$$p(\mathbf{y}^T|D_N = 1,\boldsymbol{\phi},\mathbf{Q}) = \frac{f(\mathbf{y}^T|\boldsymbol{\phi})}{r(\boldsymbol{\phi},\mathbf{Q})}\cdot D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{y}^T). \tag{6.24}$$

shocks can also be implemented in the framework we consider.

$f(\mathbf{y}^T|\boldsymbol{\phi})$ is the joint density of the data given $\boldsymbol{\phi}$ (i.e., the likelihood function of the reduced-form VAR), which depends only on $\boldsymbol{\phi}$ and the data. The indicator function $D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{y}^T)$ is equal to one when the NR are satisfied and is equal to zero otherwise. This determines the truncation points of the likelihood. $r(\boldsymbol{\phi},\mathbf{Q})$ is the ex ante probability that the NR are satisfied. This will be a constant when there are only shock-sign or shock-rank restrictions; for example, if there are s shock-sign restrictions, $r(\boldsymbol{\phi},\mathbf{Q}) = (1/2)^s$. In contrast, when there are restrictions on the historical decomposition, this probability will depend on $\boldsymbol{\phi}$ and $\mathbf{Q}$.

Consider the case where $\boldsymbol{\phi}$ is known, which will be the case asymptotically because $\boldsymbol{\phi}$ is point-identified. When $r(\boldsymbol{\phi},\mathbf{Q})$ depends on $\mathbf{Q}$, the conditional likelihood will be maximized at the value of $\mathbf{Q}$ that minimizes $r(\boldsymbol{\phi},\mathbf{Q})$ (within the set of values of $\mathbf{Q}$ that satisfy the restrictions). The posterior based on this likelihood will therefore place higher posterior probability on values of $\mathbf{Q}$ that result in a lower ex ante probability that the restrictions are satisfied. As discussed in Section 6.2.2, this is an artefact of conditioning on a non-ancillary event, which represents a loss of information about the parameters.

We therefore advocate constructing the likelihood without conditioning on the NR holding. The unconditional likelihood (the joint distribution of the data and D_N) can be expressed as

$$\begin{aligned} p(\mathbf{y}^T, D_N = d|\boldsymbol{\phi},\mathbf{Q}) &= \left[f(\mathbf{y}^T|\boldsymbol{\phi})D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{y}^T)\right]^d \cdot \left[f(\mathbf{y}^T|\boldsymbol{\phi})\left(1-D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{y}^T)\right)\right]^{1-d} \\ &= f(\mathbf{y}^T|\boldsymbol{\phi}) \cdot \left[D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{y}^T)\right]^d \cdot \left[1-D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{y}^T)\right]^{1-d}. \end{aligned} \tag{6.25}$$

For any value of $\boldsymbol{\phi}$ such that $\mathbf{y}^T$ is compatible with the NR, there will be a set of values of $\mathbf{Q}$ that satisfy the restrictions, which depend on the data, but the value of the unconditional likelihood will be the same for all values of $\mathbf{Q}$ within this set. The conditional posterior of $\mathbf{Q}|\boldsymbol{\phi},\mathbf{y}^T$ will therefore be proportional to the conditional prior for $\mathbf{Q}|\boldsymbol{\phi}$ in these regions. Given a fixed number of NR, the likelihood will possess flat regions even with a time-series of infinite length, so posterior inference may be sensitive to the choice of conditional prior for $\mathbf{Q}$, even asymptotically (which is also the case for the conditional likelihood when the restrictions are ancillary). This motivates considering Bayesian inferential procedures that are robust to the choice of unrevisable conditional prior for $\mathbf{Q}$, which we explore in Section 6.5.2.

6.3.4 Discussion

In this section, we briefly discuss the distributional assumptions for the structural shocks and the mechanism that generates the NR.

6.3.4.1 Distributional assumptions

Practitioners may be concerned about the robustness of inference with respect to deviations from the assumption of standard normal shocks. For instance, one could worry that the periods in which the NR are imposed are 'unusual' in the sense that the structural shocks in these periods were drawn from a distribution with, say, inflated variance or fat tails. The unconditional likelihood depends on the normality assumption only through $f(\mathbf{y}^T|\boldsymbol{\phi})$. By omitting terms in $f(\mathbf{y}^T|\boldsymbol{\phi})$ corresponding to the periods in which the NR are imposed, one can conduct inference that is robust to the distributional assumption about the shocks in these particular periods. To illustrate, consider the case where NR are imposed in period k only and assume the likelihood function for $\mathbf{y}^T$ takes the form

$$\tilde{f}(\mathbf{y}^T|\boldsymbol{\phi}) = v(\{\mathbf{y}_t - \mathbf{B}\mathbf{x}_t\}_{t\neq k}|\boldsymbol{\phi})w(\mathbf{y}_k - \mathbf{B}\mathbf{x}_k), \tag{6.26}$$

where

$$v(\{\mathbf{y}_t - \mathbf{B}\mathbf{x}_t\}_{t\neq k}|\boldsymbol{\phi}) = \prod_{t\neq k}(2\pi)^{-\frac{n}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)\right) \tag{6.27}$$

and $w(\mathbf{y}_k - \mathbf{B}\mathbf{x}_k)$ is an unknown, potentially non-normal, density. Replacing $f(\mathbf{y}^T|\boldsymbol{\phi})$ in Equation (6.25) with $v(\{\mathbf{y}_t - \mathbf{B}\mathbf{x}_t\}_{t\neq k}|\boldsymbol{\phi})$ yields an 'unconditional partial likelihood' that does not depend on the distribution of $\boldsymbol{\varepsilon}_k$, but that is still truncated by the NR. This would potentially result in a loss of information relative to a likelihood that correctly specifies the distribution of the shocks in period k. However, when NR are imposed in only a few periods, this loss of information is likely to be small. In contrast, the conditional likelihood approach cannot leave fully unspecified the distribution of the restricted structural shocks, because computing $r(\boldsymbol{\phi}, \mathbf{Q})$ requires specifying this distribution.

Concerns about heteroscedasticity or non-normality may also be alleviated by recognizing that the distributional assumption will become irrelevant asymptotically. The set of values of $\mathbf{Q}$ with non-zero unconditional likelihood depends only on $\boldsymbol{\phi}$, which summarizes the second moments of the data, and the realization of the data in the periods in which the NR are imposed. Under regularity assumptions, the likelihood (and thus the posterior) of $\boldsymbol{\phi}$ will converge to a point at the true value of $\boldsymbol{\phi}$ asymptotically regardless of whether the true data-generating process is a VAR with homoscedastic normal shocks.[9] The set of values of $\mathbf{Q}$ with non-zero likelihood will therefore converge asymptotically to the same set regardless of whether the distributional assumption is correct.

6.3.4.2 Mechanism generating NR

Note that we do not explicitly model the mechanism responsible for revealing the information underlying the NR (i.e., whether $D_N = 1$ or $D_N = 0$) or the mechanism determining the periods in which this information is revealed (e.g., the identity of k in examples above), which is consistent with the papers that impose these restrictions. If the revelation of this information depends on the data, the likelihood will be misspecified. The exact implications of this misspecification for estimation or inference will depend on assumptions about the mechanism revealing the narrative information. Exploring the consequences of such misspecification may be an interesting area for further work. In the bivariate example of Section 6.2, if the identity of k is randomly determined independently of $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T$, we can interpret the current analysis conditional on k.

6.4 Identification Under NR

This section formally analyzes identification in the SVAR under NR. Section 6.4.1 considers whether NR are point- or set-identifying in a frequentist sense. Sec-

[9] See Plagborg-Møller (2019) for a discussion of this point in the context of a structural VMA model.

tion 6.4.2 introduces the notion of a 'conditional identified set', which extends the standard notion of an identified set to the setting where the mapping from reduced-form to structural parameters depends on the realization of the data. This provides an interpretation of the mapping induced by the NR. Additionally, we make use of this object when showing the frequentist validity of our robust Bayesian procedure in Section 6.6.

6.4.1 Point-identification under NR

Denoting the true parameter value by $(\boldsymbol{\phi}_0, \mathbf{Q}_0)$, point-identification for the parametric model (6.25) requires that there is no other parameter value $(\boldsymbol{\phi}, \mathbf{Q}) \neq (\boldsymbol{\phi}_0, \mathbf{Q}_0)$ that is observationally equivalent to $(\boldsymbol{\phi}_0, \mathbf{Q}_0)$.[10]

To assess the existence or non-existence of observationally equivalent parameter points, we analyze a statistical distance between $p(\mathbf{y}^T, D_N = d|\boldsymbol{\phi}, \mathbf{Q})$ and $p(\mathbf{y}^T, D_N = d|\boldsymbol{\phi}_0, \mathbf{Q}_0)$ that metrizes observation equivalence. Specifically, in the current setting where the support of the distribution of observables can depend on the parameters, it is convenient to work with the Hellinger distance:

$$\begin{aligned} HD(\boldsymbol{\phi}, \mathbf{Q}) &\equiv \left(\sum_{d=0,1} \int_{\mathbf{Y}} \left(p^{1/2}(\mathbf{y}^T, D_N = d|\boldsymbol{\phi}, \mathbf{Q}) - p^{1/2}(\mathbf{y}^T, D_N = d|\boldsymbol{\phi}_0, \mathbf{Q}_0) \right)^2 d\mathbf{y}^T \right)^{\frac{1}{2}} \\ &= \sqrt{2}\,(1 - \mathscr{H}(\boldsymbol{\phi}, \mathbf{Q}))^{\frac{1}{2}}, \quad \text{where} \\ \mathscr{H}(\boldsymbol{\phi}, \mathbf{Q}) &\equiv \sum_{d=0,1} \int_{\mathbf{Y}} p^{1/2}(\mathbf{y}^T, D_N = d|\boldsymbol{\phi}, \mathbf{Q}) \cdot p^{1/2}(\mathbf{y}^T, D_N = d|\boldsymbol{\phi}_0, \mathbf{Q}_0) d\mathbf{y}^T, \end{aligned} \tag{6.28}$$

and $\mathbf{Y}$ is the sample space for $\mathbf{Y}^T$. As is known in the literature on minimum distance estimation (see, for example, Basu, Shioya and Park (2011)), $(\boldsymbol{\phi}, \mathbf{Q})$ and $(\boldsymbol{\phi}_0, \mathbf{Q}_0)$ are observationally equivalent if and only if $HD(\boldsymbol{\phi}, \mathbf{Q}) = 0$ or, equivalently, $\mathscr{H}(\boldsymbol{\phi}, \mathbf{Q}) = 1$.

[10] $(\boldsymbol{\phi}, \mathbf{Q}) \neq (\boldsymbol{\phi}_0, \mathbf{Q}_0)$ is observationally equivalent to $(\boldsymbol{\phi}_0, \mathbf{Q}_0)$ if $p(\mathbf{Y}^T, D_N = d|\boldsymbol{\phi}, \mathbf{Q}) = p(\mathbf{Y}^T, D_N = d|\boldsymbol{\phi}_0, \mathbf{Q}_0)$ holds for all $\mathbf{Y}^T$ and $d \in \{0, 1\}$.

We similarly define the Hellinger distance for the conditional likelihood as

$$
\begin{aligned}
&HD_c(\boldsymbol{\phi},\mathbf{Q}) \equiv \sqrt{2}\,(1-\mathscr{H}_c(\boldsymbol{\phi},\mathbf{Q}))^{\frac{1}{2}}, \quad \text{where} \\
&\mathscr{H}_c(\boldsymbol{\phi},\mathbf{Q}) \equiv \left(\int_{\mathbf{Y}} p^{1/2}(\mathbf{y}^T|D_N=1,\boldsymbol{\phi},\mathbf{Q})\cdot p^{1/2}(\mathbf{y}^T|D_N=1,\boldsymbol{\phi}_0,\mathbf{Q}_0)d\mathbf{y}^T\right)^{\frac{1}{2}}.
\end{aligned}
\tag{6.29}
$$

The next proposition analyzes the conditions for $\mathscr{H}(\boldsymbol{\phi},\mathbf{Q}) = 1$ and $\mathscr{H}_c(\boldsymbol{\phi},\mathbf{Q}) = 1$, and shows that observational equivalence of $(\boldsymbol{\phi},\mathbf{Q})$ and $(\boldsymbol{\phi}_0,\mathbf{Q}_0)$ boils down to geometric equivalence of the set of reduced-form VAR innovations satisfying the NR.

Proposition 6.4.1. *Let $(\boldsymbol{\phi}_0,\mathbf{Q}_0)$ be the true parameter value and let $\mathbf{U}\equiv\mathbf{U}(\mathbf{y}^T;\boldsymbol{\phi}) = (\mathbf{u}_1',\ldots,\mathbf{u}_T')'$ collect the reduced-form VAR innovations. Define*

$$
\mathscr{Q}^* \equiv \left\{ \begin{array}{c} \mathbf{Q}\in\mathscr{O}(n) : \{\mathbf{U} : N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}\} = \{\mathbf{U} : N(\boldsymbol{\phi}_0,\mathbf{Q}_0,\mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}\} \\ \textit{up to } f(\mathbf{Y}^T|\boldsymbol{\phi}_0)\textit{-null set}, \ \mathrm{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times 1} \end{array} \right\}.
$$

The unconditional likelihood model (6.25) and the conditional likelihood model (6.24) are globally identified (i.e., there are no observationally equivalent parameter points to $(\boldsymbol{\phi}_0,\mathbf{Q}_0)$) if and only if $\mathscr{Q}^*$ is a singleton. If the parameter of interest is an impulse response to the jth structural shock, $\eta_{i,j,h}(\boldsymbol{\phi},\mathbf{Q})$, as defined in (6.18), then $\eta_{i,j,h}(\boldsymbol{\phi},\mathbf{Q})$ is point-identified if the projection of $\mathscr{Q}^*$ onto its jth column vector is a singleton.

Proof. See Appendix D.2. □

This proposition provides a necessary and sufficient condition for global identification of SVARs by NR. As shown in the proof in Appendix D.2, $\mathscr{Q}^*$ defined in this proposition corresponds to the observationally equivalent $\mathbf{Q}$ matrices given $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, but, importantly, it does not correspond to any flat region of the observed likelihood (the conditional identified set in Definition 6.4.1 below).

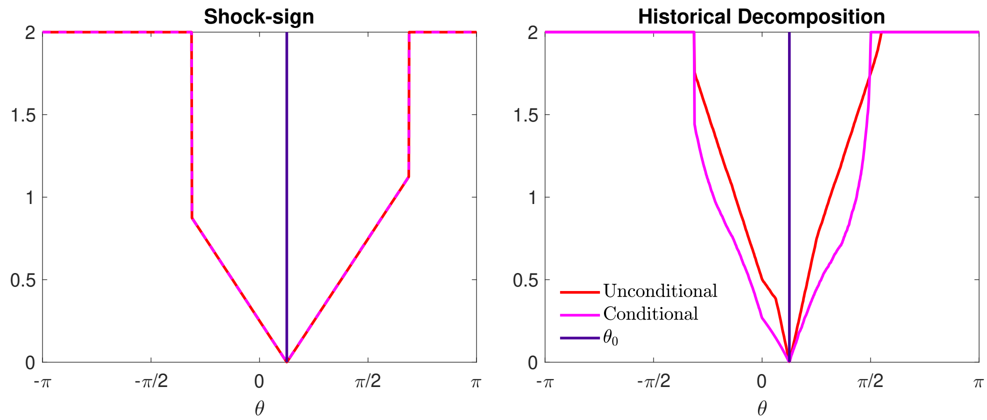
To illustrate this point, consider the simple bivariate example of Section 6.2 with the NR (6.4), where $\mathbf{y}_t$ itself is the reduced-form error, so $\mathbf{U}$ in Proposition

6.4.1 can be set to $\mathbf{y}_k$. Given $\boldsymbol{\phi}$, the set of $\mathbf{y}_k \in \mathbb{R}^2$ satisfying the NR is the half-space given by

$$\left\{ \mathbf{y}_k \in \mathbb{R}^2 : (\sigma_{11}\sigma_{22})^{-1} \left(\sigma_{22}\cos\theta - \sigma_{21}\sin\theta, \quad \sigma_{11}\sin\theta \right) \mathbf{y}_k \geq 0 \right\}. \tag{6.30}$$

The condition for point-identification shown in Proposition 6.4.1 is satisfied if no $\theta' \neq \theta$ can generate the half-space of $\mathbf{y}_k$ identical to (6.30). Such θ' cannot exist, since a half-space passing through the origin $(a_1, a_2)\mathbf{y}_k \geq 0$ can be indexed uniquely by the slope a_1/a_2 and (6.30) implies the slope $\sigma_{11}^{-1}(\sigma_{22}(\tan\theta)^{-1} - \sigma_{21})$ is a bijective map of θ on a constrained domain due to the sign normalization. Figure 6.3 plots the squared Hellinger distances in this bivariate example under the shock-sign restriction (6.4) and the historical decomposition restriction. For both the conditional and unconditional likelihood, the squared Hellinger distances are minimized uniquely at the true θ, which is consistent with our point-identification claim for θ.[11]

Figure 6.3: Squared Hellinger Distance



Notes: $T = 3$ and $\boldsymbol{\phi}$ is known; Hellinger distances are approximated using Monte Carlo.

Proposition 6.4.1 also provides conditions under which $(\boldsymbol{\phi}, \mathbf{Q})$ is not globally identified, but a particular impulse response is. To give an example of this, consider an SVAR with $n > 2$ and with a shock-sign restriction on the first shock in period k.

[11] Under the restriction on the historical decomposition, a notable difference between the conditional and unconditional likelihood cases is the slope of the squared Hellinger distance around the minimum. The squared Hellinger distance of the unconditional likelihood yields a steeper slope than the conditional likelihood. This indicates the loss of information for θ in the conditional likelihood due to conditioning on the non-ancillary event.

Given $\boldsymbol{\phi}$, the set of $\mathbf{u}_k \in \mathbb{R}^n$ satisfying the NR is a half-space defined by $\mathbf{q}_1'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_k \geq 0$. The set of values of $\mathbf{u}_k$ satisfying this inequality is indexed uniquely by $\mathbf{q}_1$ given $\boldsymbol{\Sigma}_{tr}$ at its true value, so there are no values of $\mathbf{Q}$ that are observationally equivalent to $\mathbf{Q}_0$ with $\mathbf{q}_1 \neq \mathbf{Q}_0\mathbf{e}_{1,n}$. Any value for the remaining $n-1$ columns of $\mathbf{Q}$ such that they are orthogonal to $\mathbf{Q}_0\mathbf{e}_{1,n}$ will generate the same half-space for $\mathbf{u}_k$, so $\mathscr{Q}^*$ is not a singleton and the SVAR is not globally identified. However, the projection of $\mathscr{Q}^*$ onto its first column is a singleton, so $\eta_{i,1,h}(\boldsymbol{\phi},\mathbf{Q})$ is globally identified.

Although a single NR can deliver global identification in the frequentist sense, the practical implication of this theoretical claim is not obvious. The observed unconditional likelihood is almost always flat at the maximum, so we cannot obtain a unique maximum likelihood estimator for the structural parameter. As a result, the standard asymptotic approximation of the sampling distribution of the maximum likelihood estimator is not applicable. The SVAR model with NR possesses features of set-identified models from the Bayesian standpoint (i.e., flat regions of the likelihood). However, strictly speaking, it can be classified as a globally identified model in the frequentist sense when the condition of Proposition 6.4.1 holds.

6.4.2 Conditional identified set

It is well-known that traditional sign restrictions deliver set-identification of $\mathbf{Q}$ (or, equivalently, the structural parameters). Given the reduced-form parameter $\boldsymbol{\phi}$ – which is point-identified – there are multiple observationally equivalent values of $\mathbf{Q}$, in the sense that there exists $\mathbf{Q}$ and $\tilde{\mathbf{Q}} \neq \mathbf{Q}$ such that $p(\mathbf{y}^T|\boldsymbol{\phi},\mathbf{Q}) = p(\mathbf{y}^T|\boldsymbol{\phi},\tilde{\mathbf{Q}})$ for every $\mathbf{y}^T$ in the sample space. The identified set for $\mathbf{Q}$ given $\boldsymbol{\phi}$ contains all such observationally equivalent parameter points, and is defined as

$$\mathscr{Q}(\boldsymbol{\phi}|S) = \left\{\mathbf{Q} \in \mathscr{O}(n) : S(\boldsymbol{\phi},\mathbf{Q}) \geq \mathbf{0}_{\tilde{s}\times 1}, \mathrm{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times 1}\right\}. \tag{6.31}$$

The identified set is a set-valued map only of $\boldsymbol{\phi}$, which carries all the information about $\mathbf{Q}$ contained in the data.

The complication in applying this definition of the identified set in SVARs when there are NR is that the reduced-form VAR parameters no longer represent all information about $\mathbf{Q}$ contained in the data; by truncating the likelihood, the realizations of the data entering the NR contain additional information about $\mathbf{Q}$. To address this, we introduce a refinement of the definition of an identified set.

Definition 6.4.1. Let $N \equiv N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{y}^T) \geq \mathbf{0}_{s\times 1}$ represent a set of NR in terms of the parameters and the data.

(i) The **conditional identified set for Q under NR** is

$$\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N) = \{\mathbf{Q} \in \mathscr{O}(n) : N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{y}^T) \geq \mathbf{0}_{s\times 1}\}. \tag{6.32}$$

The conditional identified set for the impulse response $\eta = \eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q})$ under NR is defined by projecting $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ via $\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q})$:

$$CIS_\eta(\boldsymbol{\phi}|\mathbf{y}^T, N) = \{\eta_{i,j,h}(\boldsymbol{\phi}, \mathbf{Q}) : \mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)\}. \tag{6.33}$$

(ii) Let $\mathbf{s} : \mathbf{Y} \to \mathbb{R}^S$ be a statistic. We call $\mathbf{s}(\mathbf{Y}^T)$ a **sufficient statistic for the conditional identified set** $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ if the conditional identified set for $\mathbf{Q}$ depends on the sample $\mathbf{y}^T$ through $\mathbf{s}(\mathbf{y}^T)$; i.e., there exists $\tilde{\mathscr{Q}}(\boldsymbol{\phi}|\cdot, N)$ such that

$$\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N) = \tilde{\mathscr{Q}}(\boldsymbol{\phi}|\mathbf{s}(\mathbf{y}^T), N) \tag{6.34}$$

holds for all $\boldsymbol{\phi} \in \boldsymbol{\Phi}$ and $\mathbf{y}^T \in \mathbf{Y}$.

Unlike the standard identified set $\mathscr{Q}(\boldsymbol{\phi}|S)$, the conditional identified set $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ depends on the sample $\mathbf{y}^T$ because of the aforementioned data-dependent support of the likelihood. In terms of the observed likelihood, however, they share the property that the likelihood is flat on the (conditional) identified set. Hence, given the sample $\mathbf{y}^T$ and the reduced-form parameters $\boldsymbol{\phi}$, all values of $\mathbf{Q}$ in $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ fit the data equally well and, in this particular sense, they are observationally equivalent.

When the NR concern shocks in only a subset of the time periods in the data, the conditional identified set under these NR depends on the sample only through a few observations entering the NR. The sufficient statistics $\mathbf{s}(\mathbf{y}^T)$ defined in Definition 6.4.1(ii) represent such observations. For instance, in the toy example of Section 6.2.1, the conditional identified set depends only on the observations in period k, so $\mathbf{s}(\mathbf{y}^T) = \mathbf{y}_k$. If we extend the example of Section 6.2.1 to the SVAR(p), the shock-sign restriction in Equation (6.4) can be expressed as

$$\varepsilon_{1k} = \mathbf{e}'_{1,2}\mathbf{A}_0\mathbf{u}_k = \mathbf{e}'_{1,2}\mathbf{Q}'\mathbf{\Sigma}_{tr}^{-1}(\mathbf{y}_k - \mathbf{B}\mathbf{x}_k) \geq 0. \tag{6.35}$$

Hence, the conditional identified set $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ depends on the data only through $(\mathbf{y}'_k, \mathbf{x}'_k)' = (\mathbf{y}'_k, \mathbf{y}'_{k-1}, \cdots, \mathbf{y}'_{k-p})'$, so we can set $\mathbf{s}(\mathbf{y}^T) = (\mathbf{y}'_k, \mathbf{y}'_{k-1}, \cdots, \mathbf{y}'_{k-p})'$.

If the conditional distribution of $\mathbf{Y}^T$ given $\mathbf{s}(\mathbf{Y}^T) = \mathbf{s}(\mathbf{y}^T)$ is nondegenerate, we can consider a frequentist experiment (repeated sampling of $\mathbf{Y}^T$) conditional on the sufficient statistics set to the observed value. In this conditional experiment, we can view the conditional identified set $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ as the standard identified set in set-identified models, since it no longer depends on the data in the conditional experiment where $\mathbf{s}(\mathbf{y}^T)$ is fixed. This is the reason that we refer to $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ as the conditional identified set. In Section 6.6 below, we show the frequentist validity of the robust-Bayes credible region by establishing conditional coverage of the conditional identified set for an impulse response.

The conditional identified set resembles the finite-sample identified set introduced by Rosen and Ura (2020) in the context of maximum score estimation (Manski (1975, 1985)), where their finite-sample identified set corresponds to the plateau of the population maximum score objective function in the conditional frequentist experiment given the regressors. In particular, if we impose only the shock-sign restrictions, the construction of the conditional identified set $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{y}^T, N)$ coincides with the construction of the finite-sample identified set for the scale-normalized coefficients given knowledge of the true data generating processes, as they both solve the system of inequalities of the form (6.4) or (6.35).[12] Despite these common geometric features, there are several differences between SVAR estimation under NR

[12] See also Komarova (2013) for the construction of the identified set for the maximum score coefficients with discrete regressors.

and maximum score estimation. First, the SVAR under NR is a likelihood-based parametric model, while maximum score estimation is a semiparametric binary regression without a likelihood. Second, NR directly trim the support of the sample objective function (i.e., likelihood) by the intersection of inequalities, while the maximum score objective function counts the number of inequalities satisfied in the sample. Third, the number of NR depends on the researcher's choice of identifying restrictions, while the number of inequalities in maximum score estimation is driven by the support points of the regressors observed in the sample.

6.5 Posterior Inference Under NR

This section presents approaches to conducting posterior inference in SVARs under NR. Section 6.5.1 discusses how to modify the standard Bayesian approach in AR18 to use the unconditional likelihood rather than the conditional likelihood. Section 6.5.2 explains how to conduct robust Bayesian inference under NR, which further addresses the issue of posterior sensitivity due to the flat unconditional likelihood. Section 6.5.3 describes how to numerically implement the robust Bayesian procedure.

6.5.1 Standard Bayesian inference

AR18 propose an algorithm for drawing from the uniform-normal-inverse-Wishart posterior of $(\boldsymbol{\phi}, \mathbf{Q})$ given a set of traditional sign restrictions and NR. This is the posterior induced by a normal-inverse-Wishart prior over $\boldsymbol{\phi}$ and an unconditionally uniform prior over $\mathbf{Q}$. The algorithm proceeds by drawing $\boldsymbol{\phi}$ from a normal-inverse-Wishart distribution and $\mathbf{Q}$ from a uniform distribution over $\mathscr{O}(n)$, and checking whether the restrictions are satisfied. If the restrictions are not satisfied, the joint draw is discarded and another draw is made. If the restrictions are satisfied, the ex ante probability that the NR are satisfied at the drawn parameter values is approximated via Monte Carlo simulation. Once the desired number of draws are obtained satisfying the restrictions, the draws are resampled with replacement using as importance weights the inverse of the probability that the NR are satisfied.[13]

This algorithm essentially draws from the posterior under the unconditional likelihood and then uses importance sampling to transform these draws into draws from the posterior given the conditional likelihood. To draw from the uniform-normal-inverse-Wishart posterior using the unconditional likelihood to construct the posterior, one therefore simply needs to omit the importance-sampling step from this algorithm. Approximating the probability used to construct the importance weights requires Monte Carlo integration, which can be computationally expensive, particularly when the NR constrain the structural shocks in multiple periods. Omitting the importance-sampling step can therefore ease the computational burden of drawing from the posterior. However, as discussed above, standard Bayesian inference under the unconditional likelihood may be sensitive to the choice of conditional prior for $\mathbf{Q}|\boldsymbol{\phi}$, because the likelihood possesses flat regions.

By rejecting draws that do not satisfy the restrictions, the algorithm described above places more weight on draws of $\boldsymbol{\phi}$ that are less likely to satisfy the restrictions under the uniform distribution over $\mathscr{O}(n)$. As discussed in Uhlig (2017), one may instead prefer to use a prior that is *conditionally* uniform over $\mathbf{Q}|\boldsymbol{\phi}$. To draw from the posterior of $(\boldsymbol{\phi},\mathbf{Q})$ under the unconditional likelihood given an arbitrary prior over $\boldsymbol{\phi}$ and a conditionally uniform prior over $\mathbf{Q}|\boldsymbol{\phi}$, one can repeat Step 2 of Algorithm 1 in Section 6.5.3.

6.5.2 Robust Bayesian inference

This section explains how to conduct robust Bayesian inference about a scalar-valued function of the structural parameters under NR and traditional sign restrictions. The approach can be viewed as performing global sensitivity analysis to assess whether posterior conclusions are robust to the choice of prior on the flat regions of the likelihood. We assume that the object of interest is a particular impulse response η, although the discussion in this section also applies to any other scalar-valued function of the structural parameters, such as the forecast error vari-

[13] Based on the results in Arias et al. (2018), AR18 argue that their algorithm draws from a normal-generalized-normal posterior over the SVAR's structural parameters $(\mathbf{A}_0,\mathbf{A}_+)$ induced by a conjugate normal-generalized-normal prior, conditional on the restrictions.

ance decomposition or the historical decomposition.

Let $\pi_{\boldsymbol{\phi}}$ be a prior over the reduced-form parameter $\boldsymbol{\phi} \in \boldsymbol{\Phi}$, where $\boldsymbol{\Phi}$ is the space of reduced-form parameters such that $\mathscr{Q}(\boldsymbol{\phi}|S)$ is non-empty. A joint prior for $(\boldsymbol{\phi}, \mathbf{Q}) \in \boldsymbol{\Phi} \times \mathscr{O}(n)$ can be written as $\pi_{\boldsymbol{\phi},\mathbf{Q}} = \pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_{\boldsymbol{\phi}}$, where $\pi_{\mathbf{Q}|\boldsymbol{\phi}}$ is supported only on $\mathscr{Q}(\boldsymbol{\phi}|S)$. When there are only traditional identifying restrictions, $\pi_{\mathbf{Q}|\boldsymbol{\phi}}$ is not updated by the data, because the likelihood function is not a function of $\mathbf{Q}$. Posterior inference may therefore be sensitive to the choice of conditional prior, even asymptotically. As discussed above, a similar issue arises under NR. The difference under NR is that $\pi_{\mathbf{Q}|\boldsymbol{\phi}}$ is updated by the data through the truncation points of the unconditional likelihood. However, at each value of $\boldsymbol{\phi}$, the unconditional likelihood is flat over the set of values of $\mathbf{Q}$ satisfying the NR. Consequently, the conditional posterior for $\mathbf{Q}|\boldsymbol{\phi}, \mathbf{Y}^T$ is proportional to the conditional prior for $\mathbf{Q}|\boldsymbol{\phi}$ at each $\boldsymbol{\phi}$ whenever the conditional identified set for $\mathbf{Q}$ given $(\boldsymbol{\phi}, \mathbf{Y}^T)$ is nonempty.

Rather than specifying a single prior for $\mathbf{Q}|\boldsymbol{\phi}$, the robust Bayesian approach of GK considers the class of all priors for $\mathbf{Q}|\boldsymbol{\phi}$ that are consistent with the traditional identifying restrictions:

$$\Pi_{\mathbf{Q}|\boldsymbol{\phi}} = \left\{\pi_{\mathbf{Q}|\boldsymbol{\phi}} : \pi_{\mathbf{Q}|\boldsymbol{\phi}}(\mathscr{Q}(\boldsymbol{\phi}|S)) = 1\right\}. \tag{6.36}$$

Notice that we cannot impose the NR using a particular conditional prior on $\mathbf{Q}|\boldsymbol{\phi}$ due to the data-dependent mapping from $\boldsymbol{\phi}$ to $\mathbf{Q}$. However, by considering all possible conditional priors for $\mathbf{Q}|\boldsymbol{\phi}$ that are consistent with the traditional identifying restrictions, we trace out all possible conditional posteriors for $\mathbf{Q}|\boldsymbol{\phi}, \mathbf{Y}^T$ that are consistent with the traditional identifying restrictions and the NR. This is because the NR truncate the unconditional likelihood function and the traditional identifying restrictions truncate the prior for $\mathbf{Q}|\boldsymbol{\phi}$, so the posterior for $\mathbf{Q}|\boldsymbol{\phi}, \mathbf{Y}^T$ is supported only on the values of $\mathbf{Q}$ that satisfy both sets of restrictions.

Given a particular prior for $(\boldsymbol{\phi}, \mathbf{Q})$ and using the unconditional likelihood, the posterior is

$$\begin{aligned}\pi_{\boldsymbol{\phi},\mathbf{Q}|\mathbf{Y}^T,D_N=1} &\propto p(\mathbf{Y}^T,D_N=1|\boldsymbol{\phi},\mathbf{Q})\pi_{\mathbf{Q}|\boldsymbol{\phi}}\pi_{\boldsymbol{\phi}} \\ &\propto f(\mathbf{Y}^T|\boldsymbol{\phi})D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T)\pi_{\boldsymbol{\phi}}\pi_{\mathbf{Q}|\boldsymbol{\phi}} \\ &\propto \pi_{\boldsymbol{\phi}|\mathbf{Y}^T}\pi_{\mathbf{Q}|\boldsymbol{\phi}}D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T). \end{aligned} \tag{6.37}$$

The final expression for the posterior makes it clear that any prior for $\mathbf{Q}|\boldsymbol{\phi}$ that is consistent with the traditional identifying restrictions is in effect further truncated by the NR (through the likelihood) once the data are realized. Generating this posterior using every prior within the class of priors for $\mathbf{Q}|\boldsymbol{\phi}$ generates a class of posteriors for $(\boldsymbol{\phi},\mathbf{Q})$:

$$\Pi_{\boldsymbol{\phi},\mathbf{Q}|\mathbf{Y}^T,D_N=1} = \left\{\pi_{\boldsymbol{\phi},\mathbf{Q}|\mathbf{Y}^T,D_N=1} = \pi_{\boldsymbol{\phi}|\mathbf{Y}^T}\pi_{\mathbf{Q}|\boldsymbol{\phi}}D_N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) : \pi_{\mathbf{Q}|\boldsymbol{\phi}} \in \Pi_{\mathbf{Q}|\boldsymbol{\phi}}\right\}. \tag{6.38}$$

Marginalizing each posterior in this class of posteriors induces a class of posteriors for η, $\Pi_{\eta|\mathbf{Y}^T,D_N=1}$. Each prior within the class of priors $\Pi_{\mathbf{Q}|\boldsymbol{\phi}}$ therefore induces a posterior for η. Associated with each of these posteriors are quantities such as the posterior mean, median and other quantiles. For example, as we consider each possible prior within $\Pi_{\mathbf{Q}|\boldsymbol{\phi}}$, we can trace out the set of all possible posterior means for η. This will always be an interval, so we can summarize this ‘set of posterior means’ by its endpoints:

$$\left[\int_{\boldsymbol{\Phi}} l(\boldsymbol{\phi},\mathbf{Y}^T)d\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}, \int_{\boldsymbol{\Phi}} u(\boldsymbol{\phi},\mathbf{Y}^T)d\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}\right], \tag{6.39}$$

where $l(\boldsymbol{\phi},\mathbf{Y}^T) = \inf\{\eta(\boldsymbol{\phi},\mathbf{Q}) : \mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)\}$, $u(\boldsymbol{\phi},\mathbf{Y}^T) = \sup\{\eta(\boldsymbol{\phi},\mathbf{Q}) : \mathbf{Q} \in \mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)\}$ and

$$\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S) = \left\{\mathscr{Q}(\boldsymbol{\phi}|S) \cap \mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N)\right\} \tag{6.40}$$

is the set of values of $\mathbf{Q}$ that are consistent with the traditional identifying restrictions and the NR. In contrast, in GK the set of posterior means is obtained by finding the infimum and supremum of $\eta(\boldsymbol{\phi},\mathbf{Q})$ over $\mathscr{Q}(\boldsymbol{\phi}|S)$ and averaging these over $\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}$. The important difference from GK is that the current set of posterior means depends on the data not only through the posterior for $\boldsymbol{\phi}$ but also through the set of admissible values of $\mathbf{Q}$ under the NR. As a result, in contrast with GK, we cannot interpret the set of posterior means (6.39) as a consistent estimator for the identified set for η (which is not well-defined, as we discussed above). Nevertheless, the set of posterior means still carries a robust Bayesian interpretation similar to GK in that it clarifies posterior results that are robust to the choice of prior on the non-updated part of the parameter space (i.e., on the flat regions of the likelihood).

As in GK, we can also report a robust credible region with credibility level α, which is the shortest interval estimate for η such that the posterior probability put on the interval is greater than or equal to α uniformly over the posteriors in $\Pi_{\eta|\mathbf{Y}^T,D_N=1}$ (see Proposition 1 of GK). One may also be interested in posterior lower and upper probabilities, which are the infimum and supremum, respectively, of the probability for a hypothesis over all posteriors in the class.

GK provide conditions under which their robust Bayesian approach has a valid frequentist interpretation, in the sense that the robust credible region is an asymptotically valid confidence set for the true identified set. For the same reason as mentioned above, however, frequentist validity of the robust credible region does not immediately extend to the NR case. We provide conditions under which the robust credible region has a valid frequentist interpretation in Section 6.6.

6.5.3 Numerical implementation of robust Bayesian approach

This section describes a general algorithm to implement our robust Bayesian procedure under NR. GK propose numerical algorithms for conducting robust Bayesian inference in SVARs identified using traditional sign and zero restrictions. Their Algorithm 1 uses a numerical optimization routine to obtain the lower and upper bounds of the identified set at each draw of $\boldsymbol{\phi}$. Obtaining the bounds via numerical optimization is not generally applicable under the class of NR considered in AR18, since the constraints on the historical decomposition are not differentiable everywhere in $\mathbf{Q}$. We therefore adapt Algorithm 2 of GK, which approximates the bounds of the identified set at each draw of $\boldsymbol{\phi}$ using Monte Carlo simulation.

Algorithm 1. *Let $N(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}$ be the set of NR and let $S(\boldsymbol{\phi},\mathbf{Q}) \geq \mathbf{0}_{\tilde{s}\times 1}$ be the set of traditional sign restrictions (excluding the sign normalization). Assume the object of interest is $\eta_{i,j^*,h} = c'_{i,h}(\boldsymbol{\phi})\mathbf{q}_{j^*}$.*

- ***Step 1**: Specify a prior for $\boldsymbol{\phi}$, $\pi_{\boldsymbol{\phi}}$, and obtain the posterior $\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}$.*

- ***Step 2**: Draw $\boldsymbol{\phi}$ from $\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}$ and check whether $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)$ is empty using the subroutine below.*

  - ***Step 2.1**: Draw an $n\times n$ matrix of independent standard normal random variables, $\mathbf{Z}$, and let $\mathbf{Z} = \tilde{\mathbf{Q}}\mathbf{R}$ be the QR decomposition of $\mathbf{Z}$.*[14]

  - ***Step 2.2**: Define*

$$\mathbf{Q} = \left[\mathrm{sgn}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\tilde{\mathbf{q}}_1)\frac{\tilde{\mathbf{q}}_1}{\|\tilde{\mathbf{q}}_1\|},\ldots,\mathrm{sgn}((\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n,n})'\tilde{\mathbf{q}}_n)\frac{\tilde{\mathbf{q}}_n}{\|\tilde{\mathbf{q}}_n\|}\right],$$

    where $\tilde{\mathbf{q}}_j$ is the jth column of $\tilde{\mathbf{Q}}$.

  - ***Step 2.3**: Check whether $\mathbf{Q}$ satisfies $\mathbf{N}(\boldsymbol{\phi},\mathbf{Q},\mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}$ and $S(\boldsymbol{\phi},\mathbf{Q}) \geq \mathbf{0}_{\tilde{s}\times 1}$. If so, retain $\mathbf{Q}$ and proceed to Step 3. Otherwise, repeat Steps 2.1 and 2.2 (up to a maximum of L times) until $\mathbf{Q}$ is obtained satisfying the restrictions. If no draws of $\mathbf{Q}$ satisfy the restrictions, approximate $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)$ as being empty and return to Step 2.*

- ***Step 3**: Repeat Steps 2.1–2.3 until K draws of $\mathbf{Q}$ are obtained. Let $\{\mathbf{Q}_k, k = 1,\ldots,K\}$ be the K draws of $\mathbf{Q}$ that satisfy the restrictions and let $\mathbf{q}_{j^*,k}$ be the j^*th column of $\mathbf{Q}_k$. Approximate $[l(\boldsymbol{\phi},\mathbf{Y}^T),u(\boldsymbol{\phi},\mathbf{Y}^T)]$ by $[\min_k \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q}_{j^*,k}, \max_k \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q}_{j^*,k}]$.*

- ***Step 4**: Repeat Steps 2–3 M times to obtain $[l(\boldsymbol{\phi}_m,\mathbf{Y}^T),u(\boldsymbol{\phi}_m,\mathbf{Y}^T)]$ for $m = 1,\ldots,M$. Approximate the set of posterior means using the sample averages of $l(\boldsymbol{\phi}_m,\mathbf{Y}^T)$ and $u(\boldsymbol{\phi}_m,\mathbf{Y}^T)$.*

[14]This is the algorithm used by Rubio-Ramírez et al. (2010) to draw from the uniform distribution over $\mathscr{O}(n)$, except that we do not normalize the diagonal elements of $\mathbf{R}$ to be positive. This is because we impose a sign normalization based on the diagonal elements of $\mathbf{A}_0 = \mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}$ in Step 2.2.

- ***Step 5**: To obtain an approximation of the smallest robust credible region with credibility $\alpha \in (0,1)$, define $d(\eta, \boldsymbol{\phi}, \mathbf{Y}^T) = \max\{|\eta - l(\boldsymbol{\phi}, \mathbf{Y}^T)|, |\eta - u(\boldsymbol{\phi}, \mathbf{Y}^T)|\}$ and let $\hat{z}_\alpha(\eta)$ be the sample α-th quantile of $\{d(\eta, \boldsymbol{\phi}_m, \mathbf{Y}^T), m = 1, ..., M\}$. An approximated smallest robust credible interval for $\eta_{i,j^*,h}$ is an interval centered at $\arg\min_\eta \hat{z}_\alpha(\eta)$ with radius $\min_\eta \hat{z}_\alpha(\eta)$.*

Algorithm 1 approximates $[l(\boldsymbol{\phi}, \mathbf{Y}^T), u(\boldsymbol{\phi}, \mathbf{Y}^T)]$ at each draw of $\boldsymbol{\phi}$ via Monte Carlo simulation. The approximated set will be too narrow given a finite number of draws of $\mathbf{Q}$, but the approximation error will vanish as the number of draws goes to infinity. The algorithm may be computationally demanding when the restrictions substantially truncate $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T, N, S)$, because many draws of $\mathbf{Q}$ from $\mathscr{O}(n)$ may be rejected at each draw of $\boldsymbol{\phi}$. However, the same draws of $\mathbf{Q}$ can be used to compute $l(\boldsymbol{\phi}, \mathbf{Y}^T)$ and $u(\boldsymbol{\phi}, \mathbf{Y}^T)$ for different objects of interest, which cuts down on computation time. For example, the same draws of $\mathbf{Q}$ can be used to compute the impulse responses of all variables to all shocks at all horizons of interest. They can also be used to compute other parameters by replacing $\eta_{i,j^*,h}$ with some other function, such as the forecast error variance decomposition, an element of $\mathbf{A}_0$, the historical decomposition or the structural shocks themselves in particular periods.[15] Step 3 is parallelizable, so reductions in computing time are possible by distributing computation across multiple processors.

6.6 Frequentist Coverage Under a Few NR

In this section, we show that the robust Bayes credible region attains asymptotically valid frequentist coverage in a setting where the number of NR is small relative to the length of the sampled periods in a sense that we make precise in the next assumption. This assumption is empirically relevant given that applications typically impose these restrictions in at most a handful of periods.

Assumption 6.6.1. (fixed-dimensional $\mathbf{s}(\mathbf{Y}^T)$): The conditional identified set under

[15] Impulse responses to a unit shock – rather than a standard-deviation shock – can be computed as in Algorithm 3 of Giacomini et al. (2022a).

NR has sufficient statistics $\mathbf{s}(\mathbf{Y}^T)$, as defined in Definition 6.4.1(ii), and the dimension of $\mathbf{s}(\mathbf{Y}^T)$ does not depend on T.

Let $(\boldsymbol{\phi}_0, \mathbf{Q}_0)$ be the true parameter values. We view the sample $\mathbf{Y}^T$ as being drawn from $p(\mathbf{Y}^T|\boldsymbol{\phi}_0)$. Let $p(\mathbf{Y}^T|\boldsymbol{\phi}_0, \mathbf{s})$ be the conditional distribution of the sample $\mathbf{Y}^T$ given the sufficient statistics for the conditional identified set $\mathbf{s} = \mathbf{s}(\mathbf{Y}^T)$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. We denote by $p(\mathbf{s}|\boldsymbol{\phi}_0)$ the distribution of the sufficient statistics $\mathbf{s}(\mathbf{Y}^T)$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. The next assumption assumes that in the conditional experiment given $\mathbf{s}(\mathbf{Y}^T)$, the sampling distribution for the maximum likelihood estimator $\hat{\boldsymbol{\phi}} \equiv \arg\max_{\boldsymbol{\phi}} p(\mathbf{Y}^T|\boldsymbol{\phi})$ centered at $\boldsymbol{\phi}_0$ and the posterior for $\boldsymbol{\phi}$ centered at $\hat{\boldsymbol{\phi}}$ asymptotically coincide.

Assumption 6.6.2. (Conditional Bernstein-von Mises property for $\boldsymbol{\phi}$): For $p(\mathbf{s}|\boldsymbol{\phi}_0)$-almost every $\mathbf{s}$ and $p(\mathbf{Y}^T|\boldsymbol{\phi}_0, \mathbf{s})$-almost every sampling sequence $\mathbf{Y}^T$, the posterior for $\sqrt{T}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})$ asymptotically coincides with the sampling distribution of $\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ with respect to $p(\mathbf{Y}^T|\boldsymbol{\phi}_0, \mathbf{s})$, as $T \to \infty$, in the sense stated in Assumption 5(i) in GK.

This is a key assumption for establishing the asymptotic frequentist validity of the robust credible region under NR. It holds, for instance, when $\mathbf{s}(\mathbf{y}^T)$ corresponds to one or a few observations in the whole sample, as we had in the toy example of Section 6.2.1. In this case, the influence of $\mathbf{s}(\mathbf{y}^T)$ vanishes in the conditional sampling distribution of $\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ as $T \to \infty$, as the latter asymptotically agrees with the asymptotically normal sampling distribution for the maximum likelihood estimator with variance-covariance matrix given by the inverse of the Fisher information matrix. By the well-known Bernstein-von Mises theorem for regular parametric models, the posterior for $\sqrt{T}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})$ asymptotically agrees with this sampling distribution.

The last assumption requires convexity and smoothness of the conditional identified set, and is analogous to Assumption 5(ii) of GK for standard set-identified models.

Assumption 6.6.3. (Almost-sure convexity and smoothness of the impulse response identified set): Let $\widetilde{CIS}_\eta(\boldsymbol{\phi}|\mathbf{s}(\mathbf{Y}^T), N)$ be the conditional identified set for η with the sufficient statistics $\mathbf{s}(\mathbf{Y}^T)$. For $p(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-almost every $\mathbf{Y}^T$, $\widetilde{CIS}_\eta(\boldsymbol{\phi}|\mathbf{s}(\mathbf{y}^T),N)$ is closed and convex, $\widetilde{CIS}_\eta(\boldsymbol{\phi}|\mathbf{s}(\mathbf{y}^T),N) = [\tilde{\boldsymbol{\ell}}(\boldsymbol{\phi},\mathbf{s}(\mathbf{Y}^T)),\tilde{\mathbf{u}}(\boldsymbol{\phi},\mathbf{s}(\mathbf{Y}^T))]$, and its lower and upper bounds are differentiable in $\boldsymbol{\phi}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ with nonzero derivatives.

Propositions D.2.1–D.2.3 in Appendix D.2 provide primitive conditions for Assumption 6.6.3 to hold in the case where there are shock-sign restrictions. Imposing Assumptions 6.6.1, 6.6.2 and 6.6.3, we obtain the following theorem.

Theorem 6.6.4. *For $\gamma \in (0,1)$, let $\widehat{C}^*_\alpha$ be the volume-minimizing robust credible region for η with credibility α,[16] which satisfies*

$$\inf_{\pi \in \Pi_{\boldsymbol{\phi},\mathbf{Q}|\mathbf{Y}^T,D_N=1}} \pi(\widehat{C}^*_\alpha) = \pi_{\boldsymbol{\phi}|\mathbf{Y}^T,D_N=1}(CIS_\eta(\boldsymbol{\phi}|\mathbf{Y}^T,N) \subset \hat{C}^*_\alpha|\mathbf{Y}^T,D_N=1) = \alpha. \tag{6.41}$$

Under Assumptions 6.6.1, 6.6.2, and 6.6.3, $\widehat{C}^*_\alpha$ attains asymptotically valid coverage for the true impulse response, η_0, conditional on $\mathbf{s}(\mathbf{Y}^T)$.

$$\begin{aligned}\liminf_{T\to\infty} P_{\mathbf{Y}^T|\mathbf{s},\boldsymbol{\phi}}(\eta_0 \in \widehat{C}^*_\alpha|\mathbf{s}(\mathbf{Y}^T),\boldsymbol{\phi}_0) \geq \\ \lim_{T\to\infty} P_{\mathbf{Y}^T|\mathbf{s},\boldsymbol{\phi}}(\widetilde{CIS}_\eta(\boldsymbol{\phi}_0|\mathbf{s}(\mathbf{Y}^T),N) \subset \widehat{C}^*_\alpha|\mathbf{s}(\mathbf{Y}^T),\boldsymbol{\phi}_0) = \alpha.\end{aligned} \tag{6.42}$$

Accordingly, $\widehat{C}^*_\alpha$ attains asymptotically valid coverage for η_0 unconditionally,

$$\liminf_{T\to\infty} P_{\mathbf{Y}^T|\boldsymbol{\phi}}(\eta_0 \in \widehat{C}^*_\alpha|\boldsymbol{\phi}_0) \geq \lim_{T\to\infty} P_{\mathbf{Y}^T|\boldsymbol{\phi}}(\widetilde{CIS}_\eta(\boldsymbol{\phi}_0|\mathbf{s}(\mathbf{Y}^T),N) \subset \widehat{C}^*_\alpha|\boldsymbol{\phi}_0) = \alpha. \tag{6.43}$$

Proof. See Appendix D.2. □

This theorem shows that the robust credible region of GK applied to the SVAR model with NR attains asymptotically valid frequentist coverage for the true impulse response as well as the conditional impulse-response identified set. Even if the

[16]The volume-minimizing robust credible region $\widehat{C}^*_\alpha$ is defined as a shortest interval among the connected intervals C_α satisfying

$$P_{\mathbf{Y}^T|\mathbf{s},\boldsymbol{\phi}}(\widetilde{CIS}_\eta(\boldsymbol{\phi}_0|\mathbf{s}(\mathbf{Y}^T),N) \subset C_\alpha|\mathbf{s}(\mathbf{Y}^T),\boldsymbol{\phi}_0) \geq \alpha.$$

See Proposition 1 in GK for a procedure to compute the volume-minimizing credible region.

point-identification condition of Proposition 6.4.1 holds for the impulse response, it is not obvious if the standard Bayesian credible region can attain frequentist coverage. This is because the Bernstein-von Mises theorem does not seem to hold for the impulse response due to the non-standard features of models with NR.

One could also consider asymptotics under an increasing number of restrictions. We conjecture that, under certain assumptions about how the NR are generated, the class of posteriors for η will converge to $\boldsymbol{\phi}_0$. Also, one can construct an asymptotically normal point estimator of the true impulse response (e.g. the posterior mean of the midpoint of the identified set). We do not focus on this case here, since the assumption that there is a fixed number of restrictions seems to be of primary interest. We leave exploration of these conjectures for further work.

6.7 Empirical Application

AR18 estimate the effects of monetary policy shocks on the US economy using a combination of sign restrictions on impulse responses and NR. The reduced-form VAR is the same as that used in Uhlig (2005). The model's endogenous variables are real GDP, the GDP deflator, a commodity price index, total reserves, non-borrowed reserves (all in natural logarithms) and the federal funds rate; see Arias, Caldara and Rubio-Ramírez (2019) for details on the variables. The data are monthly and run from January 1965 to November 2007. The VAR includes 12 lags and we include a constant.

As NR, AR18 impose that the monetary policy shock in October 1979 was positive and that it was the overwhelming contributor to the unexpected change in the federal funds rate in that month. This was the month in which the Federal Reserve markedly and unexpectedly increased the federal funds rate following the appointment of Paul Volcker as chairman of the Federal Reserve, and is widely considered to be an example of a positive monetary policy shock (e.g., Romer and Romer (1989)). The traditional sign restrictions considered in Uhlig (2005) are also imposed. Specifically, the response of the federal funds rate is restricted to be non-negative for $h = 0, 1, \ldots, 5$ and the responses of the GDP deflator, the commodity price index and nonborrowed reserves are restricted to be nonpositive for $h = 0, 1, \ldots, 5$.

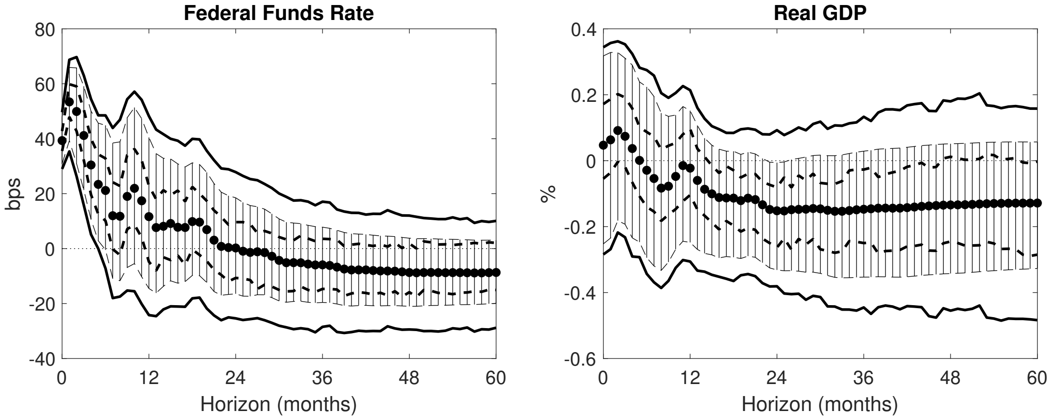
We assume a Jeffreys' (improper) prior over the reduced-form parameters, $\pi_{\boldsymbol{\phi}} = \pi_{\mathbf{B},\boldsymbol{\Sigma}} \propto |\boldsymbol{\Sigma}|^{-\frac{n+1}{2}}$, which is truncated so that the VAR is stable. The posterior for the reduced-form parameters, $\pi_{\boldsymbol{\phi}|\mathbf{Y}^T}$, is then a normal-inverse-Wishart distribution, from which it is straightforward to obtain independent draws (for example, see Del Negro and Schorfheide (2011)). We obtain 1,000 draws from the posterior of $\boldsymbol{\phi}$ such that the VAR is stable and $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T, N, S)$ is non-empty. We use Algorithm 1 with $K = 10,000$ draws of $\mathbf{Q}$ at each draw of $\boldsymbol{\phi}$ to approximate $l(\boldsymbol{\phi}, \mathbf{Y}^T)$ and $u(\boldsymbol{\phi}, \mathbf{Y}^T)$. If we cannot obtain a draw of $\mathbf{Q}$ satisfying the restrictions after 100,000 draws of $\mathbf{Q}$, we approximate $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T, N, S)$ as being empty at that draw of $\boldsymbol{\phi}$.

We explore the sensitivity of posterior inference to the choice of prior for $\mathbf{Q}|\boldsymbol{\phi}$ when the unconditional likelihood is used to construct the posterior. For brevity, we report only the impulse responses of the federal funds rate and real GDP to a positive standard-deviation monetary policy shock (Figure 6.4). As a point of comparison, we report results obtained using a conditionally uniform prior for $\mathbf{Q}|\boldsymbol{\phi}$. Under this prior, the 68 per cent highest posterior density credible intervals for the response of real GDP exclude zero at horizons greater than a year or so.[17] In contrast, the 68 per cent robust credible intervals include zero at all horizons. Under the single prior, the posterior probability that the output response is negative two years after the shock is 95 per cent. In contrast, the posterior lower probability of this event – the smallest probability over the class of posteriors generated by the class of priors – is only 54 per cent. The results suggest that posterior inference about the effect of monetary policy on output can be sensitive to the choice of (unrevisable) prior for $\mathbf{Q}|\boldsymbol{\phi}$.

AR18 also consider an alternative set of restrictions. Specifically, they impose that the monetary policy shock was: positive in April 1974, October 1979, De-

[17]The results are not directly comparable to those presented in Figure 6 of AR18. First, we present responses to a standard-deviation shock, whereas AR18 describe their responses as being to a 25 basis point shock (although, from close inspection of their Figure 6, it is evident that this normalization is not imposed correctly, because the impact response of the federal funds rate fans out around zero). Second, we use a prior for $\mathbf{Q}$ that is conditionally uniform given $\boldsymbol{\phi}$, whereas AR18 use a prior that is unconditionally uniform.

Figure 6.4: Impulse Responses to a Monetary Policy Shock



Notes: Circles and dashed lines are, respectively, posterior means and 68 per cent (pointwise) highest posterior density intervals under the uniform prior for $\mathbf{Q}|\boldsymbol{\phi}$; vertical bars are sets of posterior means and solid lines are 68 per cent (pointwise) robust credible regions obtained using Algorithm 1 with 10,000 draws from $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T, N, S)$; results are based on 1,000 draws from the posterior of $\boldsymbol{\phi}$ with nonempty $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T, N, S)$; impulse responses are to a standard-deviation shock.

cember 1988 and February 1994; negative in December 1990, October 1998, April 2001 and November 2002; and the most important contributor to the observed unexpected change in the federal funds rate in these months. The choice of these dates is based on a synthesis of information from different sources, including the chronology of monetary policy actions from Romer and Romer (1989), an updated series of the monetary policy shocks constructed using Greenbook forecasts in Romer and Romer (2004), the high-frequency monetary policy surprises from Gürkaynak, Sack and Swanson (2005), and minutes from Federal Open Markets Committee meetings. Under this extended set of restrictions, the set of posterior means and the robust credible interval are tightened noticeably, particularly at shorter horizons (Figure 6.5). The posterior lower probability of a negative output response two years after the shock is now 80 per cent, compared with 54 per cent under the October 1979 restrictions.

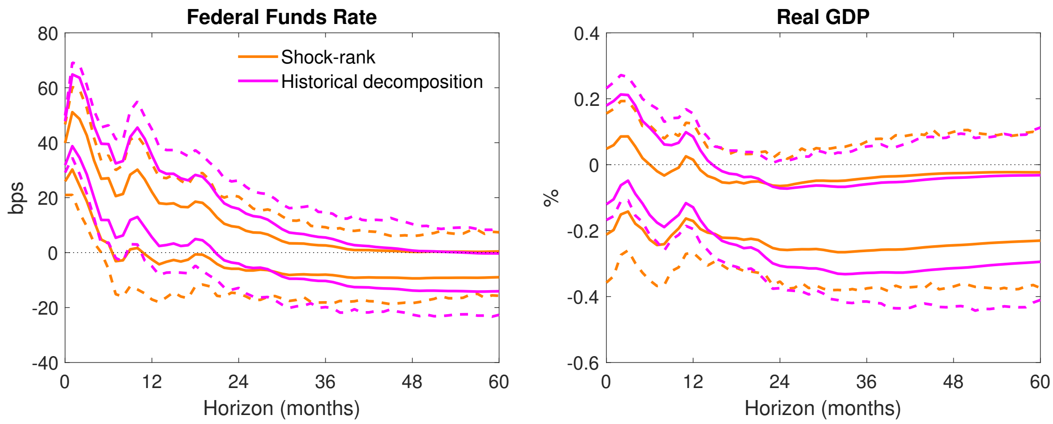
Finally, we investigate how posterior inference about the output response is affected by replacing AR18's extended set of restrictions with a shock-rank restriction. Specifically, we estimate the set of output responses that are consistent with the restriction that the monetary policy shock in October 1979 was the largest positive realization of the monetary policy shock in the sample period.[18] This restriction appears plausible given that the change in the federal funds rate in October 1979 was the largest positive change in the periods identified by AR18 as containing notable monetary policy shocks (Table 6.1). The shock-rank restriction somewhat shrinks the set of posterior means and robust credible regions relative to those obtained under the restrictions on the historical decomposition. Nevertheless, the two sets of restrictions lead to similar (robust) posterior inferences about the output response. The 68 per cent robust credible intervals include zero at all horizons under both sets of restrictions. The posterior lower probability that output falls two years after the shock is 73 per cent under the shock-rank restriction, compared with 80 per cent under the restriction on the historical decomposition.

Table 6.1: Monthly Change in Federal Funds Rate (ppt)

Oct 79	Apr 74	Dec 88	Feb 94	Dec 90	Oct 98	Apr 01	Nov 02
2.34	1.16	0.41	0.20	–0.50	–0.44	–0.51	–0.41

Source: FRED

Figure 6.5: Impulse Responses to a Monetary Policy Shock – Extended Restrictions vs Shock-rank Restriction



Notes: Solid lines represent set of posterior means and dashed lines represent 68 per cent (pointwise) robust credible regions; results are based on 1,000 draws from the posterior of $\boldsymbol{\phi}$ with nonempty $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)$; results under shock-rank restriction are obtained using Algorithm D.1; results under restrictions on the historical decomposition are obtained using Algorithm 1 with 1,000 draws from $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)$; impulse responses are to a standard-deviation shock.

[18] The large number of inequality constraints and tight conditional identified set induced by the shock-rank restriction poses computational challenges when using Algorithm 1. We therefore use the algorithm proposed in Amir-Ahmadi and Drautzburg (2021) to obtain these results.

In general, $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)$ may be empty at particular values of $\boldsymbol{\phi}$. The proportion of draws of $\boldsymbol{\phi}$ where $\mathscr{Q}(\boldsymbol{\phi}|\mathbf{Y}^T,N,S)$ is empty can therefore be used to assess the plausibility of the restrictions (see GK). Under the October 1979 restrictions, the posterior plausibility of the restrictions is one (i.e., every draw of $\boldsymbol{\phi}$ has a nonempty conditional identified set). In contrast, the posterior plausibility under AR18's extended set of restrictions is 53 per cent, while it is only 17 per cent under the shock-rank restriction.

6.8 Conclusion

Directly restricting the values of structural shocks to be consistent with historical narratives offers a potentially useful approach to disciplining SVARs, but raises novel issues related to identification and inference. These restrictions generate a set-valued mapping from the model's reduced-form parameters to its structural parameters that depends on the realization of the data entering the restrictions. This means that these restrictions do not fit neatly into the existing framework for analyzing identification in SVARs. In particular, we show that these restrictions may be point-identifying in a frequentist sense. We also highlight issues associated with existing standard Bayesian approaches to estimation and inference. Conditioning on the restrictions holding may result in the posterior placing more weight on parameters that yield a lower ex ante probability that the restrictions are satisfied. We therefore advocate using the unconditional likelihood when constructing the posterior. However, the observed unconditional likelihood will almost always possess flat regions, which implies that a component of the prior will not be updated by the data. Posterior inference may therefore be sensitive to the choice of prior. To address this, we provide robust Bayesian tools to assess or eliminate the sensitivity of posterior inference to the choice of prior. We also provide conditions under which these tools have a valid frequentist interpretation, so our approach should appeal to both Bayesians and frequentists.

While we focus on SVARs in the paper, our analysis could be extended to other settings. For example, Plagborg-Møller and Wolf (2021a) explain how to impose traditional SVAR identifying restrictions in the local projection framework under the assumption that the structural shocks are invertible. Under the assumption of invertibility, it should also be possible to impose NR within the local projection framework, but we leave a formal analysis of this problem to future research.

Appendix A

Appendix – Chapter 2

A.1 Equilibrium Conditions and Solution Method

Intermediate goods producers. There is a unit mass of intermediate goods producers indexed by $i \in [0,1]$. These firms purchase the undifferentiated good from the production firms and use it to produce a differentiated good, denoted $\tilde{Y}_{it}$, according to the production function $\tilde{Y}_{it} = Y_{it}$, where Y_{it} is the quantity of the undifferentiated good demanded by firm i. The intermediate goods producers sell their products to the representative final good producer in a monopolistically competitive market, taking the demand schedule of the final good producer (derived below) as given. Intermediate goods producers face a quadratic price-adjustment cost – paid in units of the final good – as in Rotemberg (1982). These firms discount the future using the stochastic discount factor of the representative household.

The profit-maximisation problem faced by intermediate goods producers is

$$\max_{\{P_{it}, Y_{it}, \tilde{Y}_{it}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \Lambda_{0,t} \left\{ \frac{P_{it}}{P_t} \tilde{Y}_{it} - p_t Y_{it} - \frac{\xi}{2} \left(\frac{P_{it} - P_{i,t-1}}{P_{i,t-1}} \right)^2 Y_t \right\} \tag{A.1}$$

subject to

$$\tilde{Y}_{it} = Y_{it} \tag{A.2}$$

$$\tilde{Y}_{it} = \left(\frac{P_{it}}{P_t} \right)^{-\gamma} Y_t, \tag{A.3}$$

where P_{it} is the price of the intermediate good sold by firm i, Y_t is aggregate production of the final good and γ is the elasticity of substitution between differentiated goods in final good production. The first-order condition for the problem is

$$\left\{(1-\gamma)\left(\frac{P_{it}}{P_t}\right)^{-\gamma}+\gamma p_t\left(\frac{P_{it}}{P_t}\right)^{-\gamma-1}-\xi\left(\frac{P_{it}}{P_{i,t-1}}-1\right)\frac{P_t}{P_{i,t-1}}\right\}\frac{Y_t}{P_t}$$
$$+\xi\mathbb{E}_t\left[\Lambda_{t,t+1}\left(\frac{P_{i,t+1}}{P_{it}}-1\right)\frac{P_{i,t+1}}{P_{it}^2}Y_{t+1}\right]=0. \quad \text{(A.4)}$$

I consider symmetric equilibria where all intermediate goods producers face the same initial price, which implies that they will optimally choose the same price in each period (i.e. $P_{it}=P_t$). This assumption implies that the first-order condition above simplifies to the nonlinear Phillips Curve in Equation 2.13.

Final good producer. A representative final good producer purchases differentiated goods from the intermediate goods producers and bundles them into a final good according to the production function

$$Y_t=\left(\int_0^1 \tilde{Y}_{it}^{\frac{\gamma-1}{\gamma}}di\right)^{\frac{\gamma}{\gamma-1}}. \quad \text{(A.5)}$$

It sells the final good to households (for consumption) and intermediate goods producers (to pay price-adjustment costs) in a competitive market at price P_t. The profit-maximisation problem of the final good producer is

$$\max_{\{Y_t,\{\tilde{Y}_{it}\}_{i\in[0,1]}\}} P_tY_t-\int_0^1 P_{it}\tilde{Y}_{it}di \quad \text{(A.6)}$$

subject to Equation A.5. The first-order conditions for the problem yield the demand schedule

$$\tilde{Y}_{it}=\left(\frac{P_{it}}{P_t}\right)^{-\gamma}Y_t, \quad \text{(A.7)}$$

where

$$P_t=\left(\int_0^1 P_{it}^{1-\gamma}di\right)^{\frac{1}{1-\gamma}} \quad \text{(A.8)}$$

is the price index for the final good. Final goods producers make zero profit in equilibrium.

Solution method. Given the discretisation of the state-space for idiosyncratic productivity over a k-point grid, the production firms' Bellman equation and the transition function for $\mu_t(z)$ become a system of $2k$ equations. Denote the grid points for z as z_i, $i = 1,\ldots,k$, let $P_{ij} = \Pr(z' = z_j | z = z_i)$ be the transition probabilities of the discrete-state Markov process induced by discretisation of the AR(1) process for z and let $q(z_i)$ be the probability that a potential entrant draws $z = z_i$ given the discretisation of $Q(z)$. Combining these equations with the other equilibrium conditions derived in Section 2.3 yields the following system of equations:[1]

$$V_t(z_i) = p_t z_i \left(\frac{w_t}{\nu p_t z_i}\right)^{\frac{\nu}{\nu-1}} - w_t \left(\frac{w_t}{\nu p_t z_i}\right)^{\frac{1}{\nu-1}} + \Big[\sum_{j=1}^{k} P_{ij}\mathbb{E}_t(\Lambda_{t,t+1}V_{t+1}(z_j)) \\ - \mathbb{E}_c\left(c | c \leq \sum_{j=1}^{k} P_{ij}\mathbb{E}_t(\Lambda_{t,t+1}V_{t+1}(z_j))\right)\Big] G_c\left(\sum_{j=1}^{k} P_{ij}\mathbb{E}_t(\Lambda_{t,t+1}V_{t+1}(z_j))\right) \tag{A.9}$$

$$\mu_{t+1}(z_i) = \sum_{j=1}^{k} G_c\left(\sum_{k=1}^{k} P_{jk}\mathbb{E}_t(\Lambda_{t,t+1}V_{t+1}(z_k))\right) P_{ji}\mu_t(z_j) + MG_e(V_{t+1}(z_i))q(z_i) \tag{A.10}$$

$$N_t = \sum_{i=1}^{k} \left(\frac{w_t}{\nu p_t z_i}\right)^{\frac{1}{\nu-1}} \mu_t(z_i) \tag{A.11}$$

$$w_t = \kappa_0 C_t^{\sigma} N_t^{\kappa_1} \tag{A.12}$$

$$1 = \mathbb{E}_t\left(\Lambda_{t,t+1}\frac{R_t}{\Pi_{t+1}}\right) \tag{A.13}$$

$$\frac{R_t}{R} = \left(\frac{\Pi_t}{\Pi}\right)^{\phi} \exp(\varepsilon_t^m) \tag{A.14}$$

$$(1-\gamma) + \gamma p_t - \xi(\Pi_t - 1)\Pi_t = -\xi\mathbb{E}_t\left[\Lambda_{t,t+1}(\Pi_{t+1}-1)\Pi_{t+1}\frac{Y_{t+1}}{Y_t}\right] \tag{A.15}$$

$$Y_t = \sum_{i=1}^{k} \left(\frac{w_t}{\nu p_t z_i}\right)^{\frac{\nu}{\nu-1}} \mu_t(z_i) \tag{A.16}$$

$$Y_t = C_t + \frac{\xi}{2}(\Pi_t - 1)^2 Y_t. \tag{A.17}$$

The first two equations must hold for $i = 1,\ldots,k$. Given the definition of $\Lambda_{t,t+1}$

[1] I omit the definitions of D_t and T_t since they are not necessary to solve for the equilibrium dynamics of the other variables.

and treating $V_t(z_i)$ and $\mu_t(z_i)$ as endogenous variables, this is a system of $2k+7$ equations in $2k+7$ endogenous variables, which is differentiable in the variables. Consequently, I am able to solve the model using perturbation around the stationary equilibrium (see, for example, Schmitt-Grohé and Uribe (2014)).[2] I use Dynare to solve the model using first-order perturbation (Adjemian *et al.* 2020).[3]

A.2 Representative-Firm Model

In this section, I describe the RF model used as a benchmark against which to compare the HF model developed in Section 2.3. The RF model replaces the mass of heterogeneous production firms in the HF model with a representative firm, who chooses labour input to maximise profits subject to a production function with decreasing returns to scale. The firm operates in competitive output and labour input markets. It does not pay a fixed operating cost, and there is no entry and exit. In real terms, the firm's (static) profit-maximisation problem is

$$\max_{n_t} p_t n_t^{\nu} - w_t n_t, \tag{A.18}$$

with first-order condition $w_t = \nu p_t n_t^{\nu-1}$.

The decision problems faced by the representative household, intermediate goods firms and the final goods firm are identical to those in the HF model, as is the central bank's policy rule. Log-linearising around a deterministic steady state with zero net inflation yields a system of three equations in three endogenous variables plus the exogenous monetary policy shock:

$$\hat{Y}_t - \mathbb{E}_t \hat{Y}_{t+1} = -\frac{1}{\sigma}(\hat{R}_t - \mathbb{E}_t \hat{\Pi}_{t+1}) \tag{A.19}$$

$$\hat{\Pi}_t = \frac{\gamma - 1}{\xi}\left(\sigma + \frac{\kappa_1 + 1}{\nu} - 1\right)\hat{Y}_t + \beta \mathbb{E}_t \hat{\Pi}_{t+1} \tag{A.20}$$

[2] The Blanchard-Kahn conditions are satisfied at the calibrated parameter values, so the equilibrium is unique and stable. I have verified numerically that the Blanchard-Kahn conditions are satisfied only when $\phi > 1$ given the calibrated values of the other parameters, so the well-known 'Taylor principle' holds in this model.

[3] The results are essentially identical when using second-order perturbation.

$$\hat{R}_t = \phi \hat{\Pi}_t + \varepsilon_t^m, \tag{A.21}$$

where $\hat{x}_t = \ln x_t - \ln x$ is the log-deviation of the variable x_t from its value in the deterministic steady state. The first equation is the Euler equation or IS curve, the second is the Phillips Curve and the last is the Taylor rule. Decreasing returns to scale in production ($\nu < 1$) imply that the Phillips Curve is steeper than in the textbook case with linear production technology.

A.3 Alternative Assumptions About Model Structure

This section describes alternative assumptions about the structure of the model and briefly outlines how the results change under these alternative assumptions. Unless otherwise noted, each model is recalibrated to target the same moments as in Section 2.4.1. Where appropriate, the RF model is modified to maintain comparability with the HF model.

Units of fixed costs: In the baseline model, entry and operating costs are paid in units of the final good. Alternative assumptions are that entry and operating costs are paid in units of labour or in units of the production good. Under either assumption, the entry rate increases and the exit rate decreases in response to the shock, because declines in p_t and w_t directly shift down the distribution of real fixed costs. This pattern of responses is inconsistent with the empirical evidence presented in Section 2.2. In contrast with the baseline model, where entry and exit amplify the effects of the shock, the responses of entry and exit rates in these versions of the model dampen the effects of the shock. The responses of macroeconomic variables remain essentially indistinguishable from those in the RF model.

Delay between entry and production: The baseline model assumes that entrants begin operating in the same period in which they pay the entry cost, whereas continuing firms pay the fixed operating cost in period t to continue into period $t+1$. Assuming instead that entrants pay the entry cost in period t to begin operating in period $t+1$ means that a potential entrant with idiosyncratic productivity z and entry cost e enters only if $e \leq \mathbb{E}_t(\Lambda_{t,t+1} V_{t+1}(z))$, so the real interest rate directly affects the entry decision. Excepting the delayed response of the entry rate, this alternative assumption does not substantively alter the results relative to the baseline model.

Risk-neutral firms: The baseline model assumes that production firms discount the future using the representative household's stochastic discount factor, $\Lambda_{t,t+1}$. Assuming instead that production firms are risk-neutral and discount the future at rate β means that the real interest rate no longer directly affects the value of the firm. Under this assumption, the responses of entry and exit rates are substantially smaller than in the baseline model, so there is even less amplification and propagation of the shock due to these margins.

Physical capital and aggregate frictions: This version of the model adds physical capital and other features of typical medium-scale New Keynesian models, including aggregate investment-adjustment costs, consumption habits and sticky wages.

Assume that the representative household can, in addition to saving in the nominal bond, also invest in physical capital (K_t), which is rented to the production firms in a competitive spot market at real rental rate r_{kt} and depreciates at rate δ. Investment (I_t) is undertaken subject to an investment-adjustment cost such that capital evolves according to

$$K_{t+1} = \left[1 - \frac{\tau}{2}\left(\ln\frac{I_t}{I_{t-1}}\right)^2\right] I_t + (1-\delta)K_t. \tag{A.22}$$

Also, assume that there are habits in consumption, so the household maximises

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{(C_t - hC_{t-1})^{1-\sigma} - 1}{1-\sigma} - \kappa_0 \frac{N_t^{1+\kappa_1}}{1+\kappa_1}\right). \tag{A.23}$$

Further assume that the representative household competitively supplies undifferentiated labour services to monopolistically competitive unions, who differentiate this labour and sell it to competitive labour packers subject to a quadratic wage-adjustment cost. The household owns the unions and profits generated by these unions are remitted to the household as dividends. The labour packers bundle differentiated labour services to produce undifferentiated 'final' labour services, which are sold to production firms in a competitive market.[4]

Letting λ_t and ω_t be the Lagrange multipliers on the budget constraint and the capital accumulation equation, respectively, the first-order conditions for the household's problem are:

$$w_t\lambda_t = \kappa_0 N_t \tag{A.24}$$

$$1 = \beta\mathbb{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\frac{R_t}{\Pi_{t+1}}\right) \tag{A.25}$$

$$\omega_t = \beta\mathbb{E}_t\left[\omega_{t+1}(1-\delta)+\lambda_{t+1}r_{k,t+1}\right] \tag{A.26}$$

$$\lambda_t = \omega_t\left[1-\frac{\tau}{2}\left(\ln\frac{I_t}{I_{t-1}}\right)^2-\tau\ln\frac{I_t}{I_{t-1}}\right]+\beta\tau\mathbb{E}_t\left[\omega_{t+1}\frac{I_{t+1}}{I_t}\ln\frac{I_{t+1}}{I_t}\right], \tag{A.27}$$

where $\lambda_t = (C_t - hC_{t-1})^{-\sigma} - h\beta\mathbb{E}_t(C_{t+1}-hC_t)^{-\sigma}$. The stochastic discount factor is now $\Lambda_{t,t+1} = \beta(\lambda_{t+1}/\lambda_t)^{-\sigma}$.

The structure for the labour market generates the following (nonlinear) wage Phillips Curve:

$$1-\zeta+\zeta\frac{w_t}{w_t^F}-\chi(\Pi_t^w-1)\Pi_t^w = -\chi\mathbb{E}_t\left[\Lambda_{t,t+1}(\Pi_{t+1}^w-1)\frac{\left(\Pi_{t+1}^w\right)^2}{\Pi_{t+1}}\frac{N_{t+1}}{N_t}\right], \tag{A.28}$$

where w_t is the real wage earned by households, w_t^F is the real wage paid by firms and $\Pi_t^w = \Pi_t w_t^F/w_{t-1}^F$ is the gross rate of inflation in final wages. w_t/w_t^F is real marginal cost for the labour packers, which is the inverse of the (gross) markup charged by labour unions.

Production firms operate a Cobb-Douglas production technology with decreasing returns to scale. They hire labour from the labour packers at wage rate w_t^F and rent capital from the household at rental rate r_{kt}. The production firms' Bellman equation is

$$\begin{aligned}V_t(z) = \max_{n\geq 0,k\geq 0} p_t z\left(n^\theta k^{1-\theta}\right)^\nu - w_t^F n - r_{kt}k \\ +\int\max\left\{\mathbb{E}_t\left(\Lambda_{t,t+1}V_{t+1}(z')\right)-c,0\right\}dG_c(c).\end{aligned} \tag{A.29}$$

[4]This setup follows Bayer *et al.* (2019).

The first-order conditions for this problem imply closed-form policy functions for labour demand ($n_t(z)$) and capital demand ($k_t(z)$). Using these policy functions and the threshold for exit, the Bellman equation becomes

$$V_t(z) = p_t z \left(n_t(z)^{\theta} k_t(z)^{1-\theta}\right)^{\nu} - w_t^F n_t(z) - r_{kt} k_t(z)$$
$$+ \left[\mathbb{E}_t\left(\Lambda_{t,t+1} V_{t+1}(z')\right) - \mathbb{E}_c\left(c | c < \mathbb{E}_t\left(\Lambda_{t,t+1} V_{t+1}(z')\right)\right)\right] G_c\left(\mathbb{E}_t\left(\Lambda_{t,t+1} V_{t+1}(z')\right)\right). \tag{A.30}$$

Maintaining the assumption that fixed operating and entry costs are remitted lump-sum to households, clearing of the final goods market requires

$$Y_t = C_t + I_t + \frac{\xi}{2}(\Pi_t - 1)^2 Y_t, \tag{A.31}$$

where $Y_t = \int z\left(n_t(z)^{\theta} k_t(z)^{1-\theta}\right)^{\nu} d\mu_t(z)$, clearing of the labour market requires

$$N_t = \int n_t(z) d\mu_t(z) + \frac{\chi}{2}(\Pi_t^w - 1)^2 N_t \tag{A.32}$$

and clearing of the capital market requires $K_t = \int k_t(z) d\mu_t(z)$.

I recalibrate the model to target a labour share of 0.6, which implies that $\theta = 0.8$ and I continue to set $\nu = 0.9$.[5] I assume that $\delta = 0.025$, so $r_k = 1/\beta - 1 + \delta \approx 0.035$, and that $\tau = \xi$. In a stationary equilibrium with zero net inflation in prices and final wages, $w_t / w_t^F = (\zeta - 1)/\zeta$. I assume that the markup in wage-setting is the same as in price-setting, so $\zeta = \gamma$ and the slope of the wage Phillips Curve is the same as the price Phillips Curve, so $\chi = \xi$. I set the consumption habit parameter h to 0.5.

Under these assumptions, the entry rate decreases, the exit rate increases and aggregate TFP decreases in response to a contractionary monetary policy shock. There are no substantive differences between the responses of macroeconomic variables in the HF model and a comparable RF model that includes the same frictions as the HF model. Despite the presence of sticky wages, the fact that capital is liquid

[5] This implies that the capital share is 0.15, with the remainder of output going to profits.

and there are no labour-adjustment costs means that the structure for production still induces a strong positive co-movement between the prices that enter the production firms' problem. The movements in these prices have effects on entry and exit rates that largely offset each other, so that the responses of entry and exit rates continue to be primarily driven by the direct effect of the change in the real interest rate. Quantitatively small responses of entry and exit rates – and the small size of the firms induced to exit (or not enter) – result in limited additional amplification of the shock.

Free entry: Let $\tilde{M}_t$ be the mass of *actual* entrants and assume that entrants must pay $\tilde{e}\exp(\alpha(\tilde{M}_t - \tilde{M}))$ (where $\tilde{e} > 0$ and $\alpha > 0$ are parameters) before drawing z from $Q(.)$. The assumption that the entry cost depends on the deviation of M_t from its value in the stationary equilibrium is used to avoid an unrealistically large elasticity of the entry rate to monetary policy shocks under free entry. Once an entrant has paid the entry cost and drawn z, they face the same decision problem as an incumbent firm with productivity z.[6] The value of a new entrant with idiosyncratic productivity z is therefore $V_t(z)$ (given by Equation 2.10) and the expected value of an entrant before drawing z is $V_t^e = \int V_t(z)dQ(z)$. Free entry requires that $V_t^e \leq \tilde{e}\exp(\alpha(M_t - M))$, with equality in any equilibrium with entry. I consider equilibria with entry, so the free-entry condition is $V_t^e = \tilde{e}\exp(\alpha(M_t - M))$.

The transition function for the measure of firms over idiosyncratic productivity under free entry is

$$\mu_{t+1}(B) = \int \int_{z' \in B} G_c\left(c_t^*(z)\right) dF(z'|z) d\mu_t(z) + \tilde{M}_t \int_{z \in B} dQ(z). \tag{A.33}$$

When calibrating this version of the model, as in the baseline model, I normalise the real wage to unity in the stationary equilibrium. I then solve for the parameter $\tilde{e}$ such that the free entry condition is satisfied. To solve for the mass of entrants in the stationary equilibrium, I first set $\tilde{M} = 1$ and iterate over Equation A.33 to convergence. As explained in Hopenhayn and Rogerson (1993), because $\mu(z)$ is

[6] A firm that pays the entry cost will always operate for at least one period, because the fixed operating cost does not need to be paid until the end of the period.

homogeneous of degree one in $\tilde{M}$, this gives the stationary measure of firms up to scale. The equilibrium mass of entrants can then be solved for from the labour market clearing condition given the target for the employment-to-population ratio. Consumption is obtained from the household's budget constraint and κ_0 is backed out from the labour supply condition.

Assuming that $\alpha = 15$, the entry rate decreases by about 4.5 basis points, which is similar to the point estimate of its response in the proxy SVAR. The exit rate increases, but only by about 0.8 basis points. The larger response of the entry rate relative to the exit rate is qualitatively at odds with the point estimates from the proxy SVAR. There is a quantitatively small decline in aggregate TFP and the responses of variables in the HF and RF models are again extremely similar.

Appendix B

Appendix – Chapter 3

Proof of Proposition 3.3.1: (a) Assume $\mathbf{q}_1 \in \mathbb{R}^n$ satisfies $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ and $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$, and let $\tilde{\mathbf{q}}_1 = \mathbf{K}^{-1}\mathbf{q}_1$. Since $\mathbf{K}$ is orthonormal, $\tilde{\mathbf{F}}(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 = (\mathbf{K}^{-1}\mathbf{F}(\boldsymbol{\phi})')'\mathbf{K}^{-1}\mathbf{q}_1 = \mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ and $\tilde{\mathbf{S}}(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 = (\mathbf{K}^{-1}\mathbf{S}(\boldsymbol{\phi})')'\mathbf{K}^{-1}\mathbf{q}_1 = \mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$, so $\tilde{\mathbf{q}}_1$ satisfies the restrictions in the transformed basis. It follows that

$$
\begin{aligned}
\bar{\mathbf{S}}_i(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 &= \left[\left(\mathbf{I}_n - \tilde{\mathbf{F}}(\boldsymbol{\phi})'\left(\tilde{\mathbf{F}}(\boldsymbol{\phi})\tilde{\mathbf{F}}(\boldsymbol{\phi})'\right)^{-1}\tilde{\mathbf{F}}(\boldsymbol{\phi})\right)\tilde{\mathbf{S}}_i(\boldsymbol{\phi})'\right]'\tilde{\mathbf{q}}_1 \\
&= \tilde{\mathbf{S}}_i(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 - \tilde{\mathbf{S}}_i(\boldsymbol{\phi})\tilde{\mathbf{F}}(\boldsymbol{\phi})'\left(\tilde{\mathbf{F}}(\boldsymbol{\phi})\tilde{\mathbf{F}}(\boldsymbol{\phi})'\right)^{-1}\tilde{\mathbf{F}}(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 \\
&\geq 0,
\end{aligned}
$$

for $i = 1,\ldots,s$, where the final line uses $\tilde{\mathbf{F}}(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 = \mathbf{0}_{r\times 1}$ and $\tilde{\mathbf{S}}_i(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 \geq 0$. $\tilde{\mathbf{q}}_1$ therefore satisfies the sign restrictions after projecting their coefficient vectors onto the hyperplane generated by the zero restrictions. Since $\bar{\mathbf{S}}_i(\boldsymbol{\phi})$ and $\tilde{\mathbf{q}}_1$ both lie in the hyperplane spanned by the first $n-r$ basis vectors, their last r elements are equal to zero, so $\bar{\mathbf{S}}_i(\boldsymbol{\phi})\tilde{\mathbf{q}}_1 = (\mathbf{M}\bar{\mathbf{S}}_i(\boldsymbol{\phi})')'\mathbf{M}\tilde{\mathbf{q}}_1 \geq 0$. It follows that $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$, where $\bar{\mathbf{S}}(\boldsymbol{\phi})' = \mathbf{M}(\bar{\mathbf{S}}_i(\boldsymbol{\phi})',\ldots,\bar{\mathbf{S}}_s(\boldsymbol{\phi})')$ and $\bar{\mathbf{q}}_1 = \mathbf{M}\tilde{\mathbf{q}}_1 = \mathbf{M}\mathbf{K}^{-1}\mathbf{q}_1$.

(b) Assume that $\bar{\mathbf{q}}_1 \in \mathbb{R}^{n-r}$ satisfies $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$. Given the definition of $\mathbf{M}$, $\mathbf{M}'$ is the $n\times(n-r)$ matrix such that, for an $(n-r)\times 1$ vector $\mathbf{x}$, $\mathbf{M}'\mathbf{x} = (\mathbf{x}', \mathbf{0}_{1\times r})'$. It follows that $(\mathbf{M}'\bar{\mathbf{S}}(\boldsymbol{\phi})')'\mathbf{M}'\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$. The ith column of $\mathbf{M}'\bar{\mathbf{S}}(\boldsymbol{\phi})'$ is $\bar{\mathbf{S}}_i(\boldsymbol{\phi})'$ as defined in Equation (3.8), so $\bar{\mathbf{S}}_i(\boldsymbol{\phi})\mathbf{M}'\bar{\mathbf{q}}_1 \geq 0$. From the definition of $\bar{\mathbf{S}}_i(\boldsymbol{\phi})$,

$\bar{\mathbf{S}}_i(\boldsymbol{\phi})\mathbf{M}'\bar{\mathbf{q}}_1 \geq 0$ implies that

$$\begin{aligned} & \tilde{\mathbf{S}}_i(\boldsymbol{\phi})\mathbf{M}'\bar{\mathbf{q}}_1 \geq \tilde{\mathbf{S}}_i(\boldsymbol{\phi})\tilde{\mathbf{F}}(\boldsymbol{\phi})'(\tilde{\mathbf{F}}(\boldsymbol{\phi})\tilde{\mathbf{F}}(\boldsymbol{\phi})')^{-1}\tilde{\mathbf{F}}(\boldsymbol{\phi})\mathbf{M}'\bar{\mathbf{q}}_1 \\ \Rightarrow \quad & \mathbf{S}_i(\boldsymbol{\phi})\mathbf{K}\mathbf{M}'\bar{\mathbf{q}}_1 \geq 0, \end{aligned}$$

where the last line follows from $\tilde{\mathbf{S}}_i(\boldsymbol{\phi}) = (\mathbf{K}^{-1}\mathbf{S}_i(\boldsymbol{\phi})')'$ and $\tilde{\mathbf{F}}(\boldsymbol{\phi})\mathbf{M}'\bar{\mathbf{q}}_1 = \mathbf{0}_{r\times 1}$.[1] $\mathbf{q}_1 = \mathbf{K}\mathbf{M}'\bar{\mathbf{q}}_1$ therefore satisfies $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$. Since $\tilde{\mathbf{F}}(\boldsymbol{\phi}) = (\mathbf{K}^{-1}\mathbf{F}(\boldsymbol{\phi})')'$, $\mathbf{M}'\bar{\mathbf{q}}_1$ satisfies $\mathbf{F}(\boldsymbol{\phi})\mathbf{K}\mathbf{M}\bar{\mathbf{q}}_1 = \mathbf{0}_{r\times 1}$, so $\mathbf{q}_1 = \mathbf{K}\mathbf{M}'\bar{\mathbf{q}}_1$ additionally satisfies $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$. □

Proof of Corollary 3.3.1: Assume $\bar{\mathbf{q}}_1 \in \mathbb{S}^{n-r-1}$ satisfies $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$. Then, from Proposition 3.3.1(b), $\mathbf{q}_1 = \mathbf{K}\mathbf{M}'\bar{\mathbf{q}}_1 \in \mathbb{R}^n$ satisfies $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ and $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$. Since $\bar{\mathbf{q}}_1 \in \mathbb{S}^{n-r-1}$, it has unit norm, which implies that $\mathbf{q}_1$ also has unit norm, because multiplication by $\mathbf{M}'$ adds r zeros to $\bar{\mathbf{q}}_1$ (leaving the norm unchanged) and $\mathbf{K}$ is orthonormal. $\mathbf{q}_1$ therefore lies in $\mathbb{S}^{n-1}$. Since $\mathbf{q}_1$ satisfies the identifying restrictions and lies in $\mathbb{S}^{n-1}$, it lies in $\mathscr{Q}_1(\boldsymbol{\phi}|\mathbf{F},\mathbf{S})$, which must therefore be nonempty.

Now, assume that $\mathbf{q}_1 \in \mathbb{S}^{n-1}$ satisfies $\mathbf{F}(\boldsymbol{\phi})\mathbf{q}_1 = \mathbf{0}_{r\times 1}$ and $\mathbf{S}(\boldsymbol{\phi})\mathbf{q}_1 \geq \mathbf{0}_{s\times 1}$. By Proposition 3.3.1(a), $\bar{\mathbf{q}}_1 = \mathbf{M}\mathbf{K}^{-1}\mathbf{q}_1 \in \mathbb{R}^{n-r}$ satisfies $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$. Since $\mathbf{q}_1$ has unit norm, so does $\mathbf{K}^{-1}\mathbf{q}_1$, since $\mathbf{K}^{-1}$ is orthonormal. The last r elements of $\mathbf{K}^{-1}\mathbf{q}_1$ are equal to zero, so $\bar{\mathbf{q}}_1 = \mathbf{M}\mathbf{K}^{-1}\mathbf{q}_1$ also has unit norm and thus lies in $\mathbb{S}^{n-r-1}$. Since $\bar{\mathbf{q}}_1$ satisfies $\bar{\mathbf{S}}(\boldsymbol{\phi})\bar{\mathbf{q}}_1 \geq \mathbf{0}_{s\times 1}$ and lies in $\mathbb{S}^{n-r-1}$, it lies in $\bar{\mathscr{Q}}_1(\boldsymbol{\phi}|\bar{\mathbf{S}})$, which must therefore be nonempty. □

[1] Since the last r elements of $\mathbf{M}'\bar{\mathbf{q}}_1$ are equal to zero, $\mathbf{M}'\bar{\mathbf{q}}_1$ lies in the hyperplane spanned by the first $n-r$ basis vectors. From the construction of the basis, any vector within this hyperplane lies within the null space of $\tilde{\mathbf{F}}(\boldsymbol{\phi})$, so $\tilde{\mathbf{F}}(\boldsymbol{\phi})\mathbf{M}'\bar{\mathbf{q}} = \mathbf{0}_{r\times 1}$.

Appendix C

Appendix – Chapter 5

C.1 Proofs – Strong Instruments

This appendix contains the proofs of the propositions in Section 5.3.1.

Proof of Proposition 5.3.1. Consider the first case (I), where the only restrictions are due to exogeneity of the proxies. If interest is in the impulse responses to the first shock, $j^* = 1$ by Definition 1; if interest were in the responses to the jth shock for some $j \in \{2, \dots, n-k\}$, Definition 1 would require a re-ordering of the variables such that $j^* = 1$. The exogeneity assumption requires $\mathbf{D}\mathbf{\Sigma}_{tr}\mathbf{q}_1 = \mathbf{0}_{k\times 1}$. $\mathbf{q}_1$ therefore lies in the nullspace of $\mathbf{D}\mathbf{\Sigma}_{tr}$, which, by the rank-nullity theorem, is a linear subspace of $\mathbb{R}^n$ with dimension $n-k$. Since $k < n-1$ by assumption, this subspace has dimension of at least two. The sign normalisation $(\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\mathbf{q}_1 \geq 0$ further constrains $\mathbf{q}_1$ to lie in a halfspace of $\mathbb{R}^n$. The set of feasible $\mathbf{q}_1$ is the intersection of the k-dimensional linear subspace satisfying the exogeneity restrictions, the halfspace generated by the sign normalisation and the unit sphere, which is a path-connected set. Since the impulse response is a continuous function of $\mathbf{q}_1$, the identified set is an interval and is thus convex, because the set of a continuous function with a path-connected domain is always an interval.[1]

If, instead, interest is in responses to one of the last k shocks, $j^* = n-k+1$ by Definition 1; if interest were in the responses to the jth shock for some $j \in \{n-k+2, \dots, n\}$, Definition 1 would require a re-ordering of the variables

[1]This result also follows directly from Proposition B.1(I)(i) of GK20, since $f_1 = k < n-1$.

such that $j^* = n-k+1$. For $i = 1,\dots,n-k$, $\mathbf{F}_i(\boldsymbol{\phi}) = \mathbf{D\Sigma}_{tr}$, which is a $k \times n$ matrix with rank k under the relevance assumption. $\mathbf{q}_i$, $i = 1,\dots,n-k$, lies in the nullspace of $\mathbf{D\Sigma}_{tr}$, which is of dimension $n-k$ by the rank-nullity theorem. Since the columns of an orthonormal matrix are orthogonal, $\mathbf{q}_{n-k+1}$ is orthogonal to this nullspace and so lies in the k-dimensional linear subspace of $\mathbb{R}^n$ spanned by the rows of $\mathbf{D\Sigma}_{tr}$. By assumption, $k > 1$, so this subspace has dimension of at least two. The sign normalisation $(\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{n-k+1,n})'\mathbf{q}_{n-k+1} \geq 0$ further constrains $\mathbf{q}_{n-k+1}$ to lie in a half-space of $\mathbb{R}^n$. The set of feasible $\mathbf{q}_{n-k+1}$ is the intersection of the k-dimensional linear subspace, the halfspace and the unit sphere, which is a path-connected set, and convexity of the identified set follows as above.[2]

Now consider case (II), where there are sign restrictions constraining $\mathbf{q}_{j*}$ only in addition to the exogeneity restrictions. Each of the s sign restrictions described by $\mathbf{S}_{j^*}(\boldsymbol{\phi})\mathbf{q}_{j^*} \geq \mathbf{0}_{s\times 1}$ defines a halfspace in which $\mathbf{q}_{j^*}$ must lie. The intersection of these halfspaces with the halfspace defined by the sign normalisation and the linear subspace with dimension of at least two defined by the exogeneity restrictions will still have dimension of at least two. The intersection of the resulting linear subspace with the unit sphere will therefore be a path-connected set and the impulse-response identified set will be convex.

Proof of Proposition 5.3.2. Consider case (i), where $j^* = 1$. The optimisation problem to find the upper bound of the identified set can be written as

$$u(\boldsymbol{\phi}) = \max_{\mathbf{q}\in\mathscr{S}^{n-1}} \mathbf{c}_{i,h}'(\boldsymbol{\phi})\mathbf{q} \quad \text{s.t.} \quad \mathbf{D\Sigma}_{tr}\mathbf{q} = \mathbf{0}_{k\times 1} \quad \text{and} \quad \begin{bmatrix} \mathbf{S}_1(\boldsymbol{\phi}) \\ (\mathbf{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1)\times 1}.$$

One-to-one differentiable reparameterisation of this problem using $\mathbf{x} = \mathbf{\Sigma}_{tr}\mathbf{q}$ yields the optimisation problem in Equation (2.5) of Gafarov *et al.* (2018). Differentiability of $u(\boldsymbol{\phi})$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ follows from their Theorem 2 under the assumptions that,

[2] This result does not follow from Proposition B.1 of GK20. The conditions for Proposition B.1(I)(ii) are not satisfied because $f_{j^*-1} = k \not< n-(j^*-1)$. The conditions for Proposition B.1(I)(iii) are not satisfied because there does not exist $1 \leq i^* \leq j^*-1$ such that $f_i < n-i$ for all $i = i^*+1,\dots,j^*$ and $[\mathbf{q}_1,\dots,\mathbf{q}_{i^*}]$ is exactly identified. To see this, note that the necessary condition for exact identification of $[\mathbf{q}_1,\dots,\mathbf{q}_{i^*}]$ is that $f_i = n-i$ for all $i = 1,\dots,i^*$. But $f_1 = k < n-1$, so this condition fails.

at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, the column vectors of $\left[(\mathbf{D}\boldsymbol{\Sigma}_{tr})', \mathbf{S}_1(\boldsymbol{\phi})', \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n}\right]$ are linearly independent, the set of solutions to the optimisation problem is singleton, the optimised value $u(\boldsymbol{\phi})$ is nonzero, and the number of binding sign restrictions at the optimum is less than $n-k-1$. Differentiability of $l(\boldsymbol{\phi})$ follows similarly, with $l(\boldsymbol{\phi})$ defined as the minimizer of $\mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q}$ with respect to $\mathbf{q} \in \mathscr{S}^{n-1}$ and subject to the same set of constraints.

Consider case (ii), where $j^* = n-k+1$. Let $\mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})$ be an orthonormal basis for the nullspace of $\mathbf{D}\boldsymbol{\Sigma}_{tr}$ (an $(n-k) \times n$ matrix). The optimisation problem to find the upper bound of the identified set can be written as

$$u(\boldsymbol{\phi}) = \max_{\mathbf{q} \in \mathscr{S}^{n-1}} \mathbf{c}'_{i,h}(\boldsymbol{\phi})\mathbf{q}$$

$$\text{s.t.} \quad \mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})'\mathbf{q} = \mathbf{0}_{(n-k)\times 1} \quad \text{and} \quad \begin{bmatrix} \mathbf{S}_{n-k+1}(\boldsymbol{\phi}) \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n-k+1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(s+1)\times 1}.$$

One-to-one differentiable reparameterisation of this problem using $\mathbf{x} = \boldsymbol{\Sigma}_{tr}\mathbf{q}$ yields the optimisation problem in Equation (2.5) of Gafarov *et al.* (2018) with the expanded set of equality restrictions including $\mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr})'\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{x} = \mathbf{0}_{(n-k)\times 1}$. Differentiability of $u(\boldsymbol{\phi})$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ follows from their Theorem 2 under the assumptions that, at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, the column vectors of $\left[\mathbf{N}(\mathbf{D}\boldsymbol{\Sigma}_{tr}), \mathbf{S}_{n-k+1}(\boldsymbol{\phi})', \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{n-k+1,n}\right]$ are linearly independent, the set of solutions to the optimisation problem is singleton, the optimised value $u(\boldsymbol{\phi})$ is nonzero, and the number of binding sign restrictions at the optimum is less than $n-(n-k)-1 = k-1$. Differentiability of $l(\boldsymbol{\phi})$ follows similarly.

C.2 Proofs – Weak Instruments

This appendix sets up the framework for the weak-proxy approximations of the posterior distribution and the sampling distribution of the MLE for the upper bound of the identified set, and derives formally the claims in (5.12) and (5.13).

As in Section 5.3.2, we consider the simple setting of $n = 3$ and $k = 1$, where the upper bound of the identified set $u(\boldsymbol{\phi})$ is given by (5.8). Since $u(\boldsymbol{\phi})$ depends on the reduced-form parameters only through $(\mathbf{c}, \mathbf{d})$, we express $u(\boldsymbol{\phi})$ as $u(\mathbf{c}, \mathbf{d})$. The singularity points of $u(\mathbf{c}, \mathbf{d})$ that we focus on are $\mathbf{c} \neq \mathbf{0}_{3\times 1}$ and $\mathbf{d} = \mathbf{0}_{3\times 1}$, where the weak-proxy scenario corresponds to values of $\mathbf{d}$ close to $\mathbf{0}_{3\times 1}$. We hence consider a sequence of reduced-form parameters $\{\boldsymbol{\phi}_T : T = 1, 2, \ldots\}$ along which the implied parameters $(\mathbf{c}_T, \mathbf{d}_T)$, $T = 1, 2, \ldots$, converge to $(\mathbf{c}_0, \mathbf{0}_{3\times 1})$, $\mathbf{c}_0 \neq \mathbf{0}_{3\times 1}$, as $T \to \infty$. As in the main text, we specify a drifting sequence of $\{\boldsymbol{\phi}_T\}$ that leads to

$$\begin{pmatrix} \mathbf{c}_T \\ \mathbf{d}_T \end{pmatrix} = \begin{pmatrix} \mathbf{c}_0 + \boldsymbol{\gamma}/\sqrt{T} \\ \boldsymbol{\delta}/\sqrt{T} \end{pmatrix}, \tag{B.1}$$

where $(\boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^3 \times \mathbb{R}^3$ are the localisation parameters.

Let $\hat{\mathbf{S}}_T \in \mathbb{R}^s$, $s < \infty$, $T = 1, 2, \ldots$, be a finite-dimensional vector of sufficient statistics for $\boldsymbol{\phi}$ that converges in distribution to a random vector $\hat{\mathbf{S}} \in \mathbb{R}^s$ as $T \to \infty$. Since we consider a Gaussian proxy SVAR, these sufficient statistics are the first and second sample moments of the observables. By the Skhorohod representation theorem, we can embed this sequence of sufficient statistics $\{\hat{\mathbf{S}}_T\}$ and the limiting random variables $\hat{\mathbf{S}}$ into a common probability space on which

$$\hat{\mathbf{S}}_T \to \hat{\mathbf{S}} \text{ as } T \to \infty, \text{ almost surely,} \tag{B.2}$$

holds.

Let $(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$ be the MLE of $(\mathbf{c}, \mathbf{d})$. Since the MLE depends only on the sufficient statistics $\hat{\mathbf{S}}_T$, we can embed the MLE into the probability space on which $\{\hat{\mathbf{S}}_T\}$ and $\hat{\mathbf{S}}$ are commonly defined. Hence, conditioning on the sequence of sufficient statistics $\{\hat{\mathbf{S}}_T : T = 1, 2, \ldots,\}$ pins down the constant sequence of MLEs. We assume that the (unconditional) sampling distribution of the MLEs centered at the drifting true values is asymptotically normal:

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \hat{\mathbf{c}}_T - \mathbf{c}_T \\ \hat{\mathbf{d}}_T - \mathbf{d}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \hat{\mathbf{Z}}_c \\ \hat{\mathbf{Z}}_d \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}_{6\times 1}, \begin{pmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{cd} \\ \boldsymbol{\Omega}_{cd}' & \boldsymbol{\Omega}_d \end{pmatrix} \right). \tag{B.3}$$

Following the Skhorohod representation for the sufficient statistics (B.2), we have the almost-sure convergence of the MLE to the limiting Gaussian random variables

$$\begin{pmatrix} \hat{\mathbf{Z}}_{cT} \\ \hat{\mathbf{Z}}_{dT} \end{pmatrix} \to \begin{pmatrix} \hat{\mathbf{Z}}_{c} \\ \hat{\mathbf{Z}}_{d} \end{pmatrix} \text{ as } T \to \infty, \text{ almost surely,} \tag{B.4}$$

on the common probability space. We also impose a high-level assumption of the strong consistency of the MLE for $\mathbf{c}$ in the sense of

$$\hat{\mathbf{c}}_T \to \mathbf{c}_0 \text{ as } T \to \infty, \text{ almost surely,} \tag{B.5}$$

on the same probability space.

Since the posterior distribution depends on the data only through the sufficient statistics, it suffices to consider the convergence of the posterior distribution for $u(\mathbf{c},\mathbf{d})$ conditional on the sequence of sufficient statistics $\{\hat{\mathbf{S}}_T\}$. We assume that the posterior for $(\mathbf{c},\mathbf{d})$ centered at their MLEs is asymptotically normal in the following sense. Let

$$\begin{pmatrix} \mathbf{Z}_{cT} \\ \mathbf{Z}_{dT} \end{pmatrix} \equiv \sqrt{T} \begin{pmatrix} \mathbf{c} - \hat{\mathbf{c}}_T \\ \mathbf{d} - \hat{\mathbf{d}}_T \end{pmatrix}, \tag{B.6}$$

and assume

$$\begin{pmatrix} \mathbf{Z}_{cT} \\ \mathbf{Z}_{dT} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{Z}_{c} \\ \mathbf{Z}_{d} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}_{6\times 1}, \begin{pmatrix} \mathbf{\Omega}_c & \mathbf{\Omega}_{cd} \\ \mathbf{\Omega}_{cd}' & \mathbf{\Omega}_d \end{pmatrix} \right), \tag{B.7}$$

for almost every conditioning sequence of $\{\mathbf{S}_T\}$. We assume that the asymptotic posterior variance given in (B.7) is independent of the conditioning variable $\{\hat{\mathbf{S}}_T : T = 1,2,\dots\}$ and coincides with the asymptotic variance of the MLE given in (B.3).

The asymptotic normality of the posterior (centered at the MLE with data-independent variance) holds for a wide class of regular parametric models, and its almost-sure coincidence with the asymptotic (sampling) distribution of the MLE leads to the Bernstein-von Mises Theorem. See, for instance, Schervish (1995) and DasGupta (2008) for a set of sufficient conditions for posterior asymptotic normality.

Under these assumptions, we obtain the following weak-proxy asymptotic approximation of the posterior for $u(\boldsymbol{\phi})$.

Proposition C.2.1. *Consider a drifting sequence of reduced-form parameters that satisfy (B.1) with $\mathbf{c}_0 \neq \mathbf{0}_{3\times 1}$, along which we assume that the MLE for $(\mathbf{c},\mathbf{d})$ and its posterior satisfies (B.3), (B.4), (B.5) and (B.7). Then, for almost every conditioning sequence of the sufficient statistics $\{\hat{\mathbf{S}}_T\}$, the asymptotic posterior of $u(\mathbf{c},\mathbf{d})$ is*

$$u(\mathbf{c},\mathbf{d}) \xrightarrow{d} u(\mathbf{c}_0, \boldsymbol{\delta}+\hat{\mathbf{Z}}_d+\mathbf{Z}_d) = \sqrt{\mathbf{c}_0'\left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta}+\hat{\mathbf{Z}}_d+\mathbf{Z}_d)(\boldsymbol{\delta}+\hat{\mathbf{Z}}_d+\mathbf{Z}_d)'}{\|\boldsymbol{\delta}+\hat{\mathbf{Z}}_d+\mathbf{Z}_d\|^2}\right)\mathbf{c}_0},$$

where $\hat{\mathbf{Z}}_d$ is a constant given the sampling sequence, and $\mathbf{Z}_d \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Omega}_d)$.

Proof. Since $u(\mathbf{c},\mathbf{d})$ is homogeneous of degree zero with respect to $\mathbf{d}$, we have

$$\begin{aligned} u(\mathbf{c},\mathbf{d}) &= u(\mathbf{c}, T^{1/2}\mathbf{d}) \\ &= u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, T^{1/2}\hat{\mathbf{d}}_T + \mathbf{Z}_{dT}) \\ &= u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, T^{1/2}\mathbf{d}_T + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT}) \\ &= u(\hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT}, \boldsymbol{\delta} + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT}), \end{aligned}$$

where the second equality uses (B.6), the third equality uses (B.3), and the fourth equality uses (B.1). Conditional on the sampling sequence of the sufficient statistics $\{\hat{\mathbf{S}}_T\}$, the assumptions of almost-sure convergence (B.4) and (B.5) and the posterior distributional convergence (B.7) imply

$$\begin{pmatrix} \hat{\mathbf{c}}_T + T^{-1/2}\mathbf{Z}_{cT} \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_{dT} + \mathbf{Z}_{dT} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{c}_0 \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_d + \mathbf{Z}_d \end{pmatrix},$$

as $T \to \infty$, where $(\mathbf{c}_0, \boldsymbol{\delta}, \hat{\mathbf{Z}}_d)$ are constants and $\mathbf{Z}_d$ is a random vector following $\mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Omega}_d)$. Since $u(\mathbf{c},\mathbf{d})$ is discontinuous at $\mathbf{d} = \mathbf{0}_{3\times 1}$, and $\{\boldsymbol{\delta}+\hat{\mathbf{Z}}_d+\mathbf{Z}_d = \mathbf{0}_{3\times 1}\}$ is the null event in terms of the probability law of the limiting random variables, an application of the continuous mapping theorem (see, for example, Theorem 10.8 of Kosorok (2008)) yields the conclusion. □

The next proposition gives the asymptotic sampling distribution of $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$.

Proposition C.2.2. *Consider a drifting sequence of reduced-form parameters that satisfy (B.1) with $\mathbf{c}_0 \neq \mathbf{0}_{3\times 1}$, along which we assume that the MLE of $(\mathbf{c}, \mathbf{d})$ satisfies (B.3). Then, the asymptotic distribution of $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T)$ is*

$$u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T) \xrightarrow{d} u(\mathbf{c}_0, \boldsymbol{\delta} + \hat{\mathbf{Z}}_d) = \sqrt{\mathbf{c}_0' \left(\mathbf{I}_3 - \frac{(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)(\boldsymbol{\delta} + \hat{\mathbf{Z}}_d)'}{\|\boldsymbol{\delta} + \hat{\mathbf{Z}}_d\|^2} \right) \mathbf{c}_0},$$

where $\hat{\mathbf{Z}}_d \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Omega}_d)$.

Proof. Since $u(\mathbf{c}, \mathbf{d})$ is homogeneous of degree zero with respect to $\mathbf{d}$, it holds that $u(\hat{\mathbf{c}}_T, \hat{\mathbf{d}}_T) = u(\hat{\mathbf{c}}_T, T^{1/2}\hat{\mathbf{d}}_T)$. Under the drifting sequence (B.1) and $\sqrt{T}$-asymptotic normality of the MLE (B.3),

$$\begin{pmatrix} \hat{\mathbf{c}}_T \\ T^{1/2}\hat{\mathbf{d}}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{c}_0 \\ \boldsymbol{\delta} + \hat{\mathbf{Z}}_d. \end{pmatrix}.$$

Noting that $\{\boldsymbol{\delta} + \hat{\mathbf{Z}}_d = \mathbf{0}_{3\times 1}\}$ is a null event in terms of the limiting probability law, an application of the continuous mapping theorem leads to the conclusion. $\square$

Appendix D

Appendix – Chapter 6

D.1 Bivariate Example Derivations

Set of values of θ under shock-sign restriction. This section derives analytical expressions for the set of values of θ consistent with the shock-sign restriction in the bivariate example of Section 6.2. Throughout, we assume that $\theta \in [-\pi, \pi]$.

Under the shock-sign restriction $\varepsilon_{1k} \geq 0$ and the sign normalization $\mathrm{diag}(\mathbf{A}_0) \geq \mathbf{0}_{2\times 1}$, θ is restricted to lie in the set

$$\begin{aligned}\theta \in \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \geq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\} \\ \cup \{\theta : \sigma_{21}\sin\theta \leq \sigma_{22}\cos\theta, \cos\theta \leq 0, \sigma_{22}y_{1k}\cos\theta \geq (\sigma_{21}y_{1k} - \sigma_{11}y_{2k})\sin\theta\}.\end{aligned} \tag{D.1.1}$$

Consider the case where $\sigma_{21} < 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} < 0$. Then θ is restricted to the set

$$\begin{aligned}\theta \in \left\{\theta : \tan\theta \geq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta > 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \leq \tan\theta\right\} \cup \left\{\frac{\pi}{2}\right\} \\ \cup \left\{\theta : \tan\theta \leq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta < 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \geq \tan\theta\right\}.\end{aligned} \tag{D.1.2}$$

The inequalities in the first set hold if and only if $\tan\theta \geq \max\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}$

and $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, which implies that

$$\arctan\left(\max\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right) \leq \theta < \frac{\pi}{2}. \tag{D.1.3}$$

The inequalities on the second line hold if and only if $\tan\theta \leq \min\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right\}$ and $\theta \in [-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$. Since $\sigma_{21} < 0$, $\tan\theta$ must be negative, which implies that $\theta \in (\frac{\pi}{2}, \pi]$. It follows that

$$\frac{\pi}{2} < \theta \leq \pi + \arctan\left(\min\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right). \tag{D.1.4}$$

Taking the union of (D.1.3), (D.1.4) and $\left\{\frac{\pi}{2}\right\}$ implies that

$$\theta \in \left[\arctan\left(\max\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right),\right.$$
$$\left.\pi + \arctan\left(\min\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right)\right]. \tag{D.1.5}$$

Next, consider the case where $\sigma_{21} < 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$. Then θ is restricted to the set

$$\theta \in \left\{\theta : \tan\theta \geq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta > 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \geq \tan\theta\right\}$$
$$\cup \left\{\theta : \tan\theta \leq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta < 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \leq \tan\theta\right\}. \tag{D.1.6}$$

If $y_{1k} > 0$ or if $y_{1k} < 0$ and $\frac{\sigma_{22}}{\sigma_{21}} < \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}$, the second set of inequalities is not satisfied for any θ, while the first set of inequalities is satisfied for

$$\theta \in \left[\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right)\right]. \tag{D.1.7}$$

If $y_{1k} < 0$ and $\frac{\sigma_{22}}{\sigma_{21}} > \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}$, the first set of inequalities has no solution and the second set is satisfied for

$$\theta \in \left[\pi + \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right), \pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right]. \tag{D.1.8}$$

In the case where $\sigma_{21} > 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} < 0$, θ is restricted to the set

$$\theta \in \left\{\theta : \tan\theta \leq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta > 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \leq \tan\theta\right\} \cup \left\{\theta : \tan\theta \geq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta < 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \geq \tan\theta\right\}. \quad \text{(D.1.9)}$$

If $y_{1k} > 0$ or if $y_{1k} < 0$ and $\frac{\sigma_{22}}{\sigma_{21}} > \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}$, the second set of inequalities has no solution, while the first is satisfied for

$$\theta \in \left[\arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right), \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right]. \quad \text{(D.1.10)}$$

If $y_{1k} < 0$ and $\frac{\sigma_{22}}{\sigma_{21}} < \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}$, the first set of inequalities has no solution and the second set is satisfied for

$$\theta \in \left[-\pi + \arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right), -\pi + \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right)\right]. \quad \text{(D.1.11)}$$

Finally, in the case where $\sigma_{21} > 0$ and $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$, θ is restricted to the set

$$\theta \in \left\{\theta : \tan\theta \leq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta > 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \geq \tan\theta\right\} \cup \left\{-\frac{\pi}{2}\right\} \cup \left\{\theta : \tan\theta \geq \frac{\sigma_{22}}{\sigma_{21}}, \cos\theta < 0, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}} \leq \tan\theta\right\}. \quad \text{(D.1.12)}$$

The first set of inequalities holds if and only if $\tan\theta \leq \min\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}$ and $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, which implies that

$$-\frac{\pi}{2} < \theta \leq \arctan\left(\min\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right). \quad \text{(D.1.13)}$$

The second set of inequalities holds if and only if $\tan\theta \geq \max\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}$ and $\theta \in [-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$. Since $\sigma_{21} > 0$, $\tan\theta$ must be positive, which implies that $\theta \in [-\pi, -\frac{\pi}{2})$. It follows that

$$-\pi + \arctan\left(\max\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right) \leq \theta < -\frac{\pi}{2}. \tag{D.1.14}$$

Taking the union of (D.1.13), (D.1.14) and $\left\{-\frac{\pi}{2}\right\}$ implies that

$$\theta \in \left[-\pi + \arctan\left(\max\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right),\right.$$
$$\left.\arctan\left(\min\left\{\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k} - \sigma_{11}y_{2k}}\right\}\right)\right]. \tag{D.1.15}$$

Set of values of η under shock-sign restriction. Here we derive the expression for the set of impulse responses $\eta \equiv \sigma_{11}\cos\theta$ consistent with the shock-sign restriction (i.e., (6.8) in Section 6.2).

In the absence of restrictions, the set of admissible values for the matrix of contemporaneous impulse responses is

$$\mathbf{A}_0^{-1} \in \left\{\begin{bmatrix} \sigma_{11}\cos\theta & -\sigma_{11}\sin\theta \\ \sigma_{21}\cos\theta + \sigma_{22}\sin\theta & \sigma_{22}\cos\theta - \sigma_{21}\sin\theta \end{bmatrix}\right\}$$
$$\cup \left\{\begin{bmatrix} \sigma_{11}\cos\theta & \sigma_{11}\sin\theta \\ \sigma_{21}\cos\theta + \sigma_{22}\sin\theta & \sigma_{21}\sin\theta - \sigma_{22}\cos\theta \end{bmatrix}\right\}. \tag{D.1.16}$$

Assume that $\sigma_{21} < 0$, $\sigma_{21}y_{1k} - \sigma_{11}y_{2k} > 0$ and $y_{1k} > 0$. Within the interval for θ defined in (D.1.7), η is maximized at $\theta = 0$, so $\eta_{ub} = \sigma_{11}$. The lower bound η_{lb}

occurs at one of the endpoints of the interval for θ, so it satisfies

$$\begin{aligned}
\eta_{lb} &= \min\left\{\sigma_{11}\cos\left(\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right), \sigma_{11}\cos\left(\arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right)\right)\right\} \\
&= \min\left\{\sigma_{11}\cos\left(-\arctan\left(\frac{\sigma_{22}}{\sigma_{21}}\right)\right), \sigma_{11}\cos\left(\arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right)\right)\right\} \\
&= \min\left\{\sigma_{11}\cos\left(\arctan\left(-\frac{\sigma_{22}}{\sigma_{21}}\right)\right), \sigma_{11}\cos\left(\arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right)\right)\right\} \\
&= \sigma_{11}\cos\left(\max\left\{\arctan\left(-\frac{\sigma_{22}}{\sigma_{21}}\right), \arctan\left(\frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right)\right\}\right) \\
&= \sigma_{11}\cos\left(\arctan\left(\max\left\{-\frac{\sigma_{22}}{\sigma_{21}}, \frac{\sigma_{22}y_{1k}}{\sigma_{21}y_{1k}-\sigma_{11}y_{2k}}\right\}\right)\right). \qquad \text{(D.1.17)}
\end{aligned}$$

The second line follows from the fact that $\cos(.)$ is an even function and the third line follows from the fact that $\arctan(.)$ is an odd function. The arguments entering the $\cos(.)$ functions on the third line are both in the interval $[0, \frac{\pi}{2})$, so the fourth line follows from the fact that $\cos(.)$ is a decreasing function over this domain. The final line follow from the fact that $\arctan(.)$ is an increasing function.

D.2 Omitted Proofs

Proof of Proposition 6.4.1.

Proof. $\mathscr{H}(\boldsymbol{\phi}, \mathbf{Q})$ can be written as

$$\begin{aligned}
\mathscr{H}(\boldsymbol{\phi}, \mathbf{Q}) = &\int_{\mathbf{Y}} f^{1/2}(\mathbf{y}^T|\boldsymbol{\phi}) f^{1/2}(\mathbf{y}^T|\boldsymbol{\phi}_0) \cdot D_N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{y}^T) D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{y}^T) d\mathbf{y}^T \\
&+ \int_{\mathbf{Y}} f^{1/2}(\mathbf{y}^T|\boldsymbol{\phi}) f^{1/2}(\mathbf{y}^T|\boldsymbol{\phi}_0) \cdot (1 - D_N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{y}^T))(1 - D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{y}^T)) d\mathbf{y}^T.
\end{aligned}$$

Note that the likelihood for the reduced-form parameters $f(\mathbf{y}^T|\boldsymbol{\phi})$ point-identifies $\boldsymbol{\phi}$, so $f(\cdot|\boldsymbol{\phi}) = f(\cdot|\boldsymbol{\phi}_0)$ holds only at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. Hence, we set $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ and consider $\mathscr{H}(\boldsymbol{\phi}_0, \mathbf{Q})$,

$$\mathscr{H}(\boldsymbol{\phi}_0, \mathbf{Q}) = \int_{\{\mathbf{y}^T : D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{y}^T) = D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{y}^T)\}} f(\mathbf{y}^T|\boldsymbol{\phi}_0) d\mathbf{y}^T.$$

Hence, $\mathscr{H}(\boldsymbol{\phi}_0, \mathbf{Q}) = 1$ if and only if $D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{y}^T) = D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{y}^T)$ holds $f(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-a.s. In terms of the reduced-form residuals entering the NR, the latter condition is equivalent to $\{\mathbf{U} : N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}\} = \{\mathbf{U} : N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T) \geq \mathbf{0}_{s\times 1}\}$ up to $f(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-null set. Hence, $\mathscr{Q}^*$ defined in the proposition collects observationally equivalent values of $\mathbf{Q}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ in terms of the unconditional likelihood.

Next, consider the conditional likelihood and consider

$$
\begin{aligned}
\mathscr{H}_c(\boldsymbol{\phi}_0, \mathbf{Q}) &= \frac{1}{r^{1/2}(\boldsymbol{\phi}, \mathbf{Q}) r^{1/2}(\boldsymbol{\phi}_0, \mathbf{Q}_0)} \int_{\mathbf{Y}} f(\mathbf{y}^T|\boldsymbol{\phi}_0) \cdot D_N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{y}^T) D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{y}^T) d\mathbf{y}^T \\
&= \frac{E_{\mathbf{Y}^T|\boldsymbol{\phi}_0}\left[D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{Y}^T) D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T)\right]}{r^{1/2}(\boldsymbol{\phi}, \mathbf{Q}) r^{1/2}(\boldsymbol{\phi}_0, \mathbf{Q}_0)} \\
&\leq 1,
\end{aligned}
$$

where the inequality follows by the Cauchy-Schwartz inequality, and it holds with equality if and only if $D_N(\boldsymbol{\phi}_0, \mathbf{Q}, \mathbf{Y}^T) = D_N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{Y}^T)$ holds $f(\mathbf{Y}^T|\boldsymbol{\phi}_0)$-a.s. Hence, by repeating the argument for the unconditional likelihood case, we conclude that $\mathscr{Q}^*$ consists of observationally equivalent values of $\mathbf{Q}$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ in terms of the conditional likelihood. $\square$

Proof of Theorem 6.6.4. Since $(\boldsymbol{\phi}_0, \mathbf{Q}_0)$ satisfies the imposed NR $N(\boldsymbol{\phi}_0, \mathbf{Q}_0, \mathbf{y}^T) \geq \mathbf{0}_{s\times 1}$ and the other sign restrictions (if any imposed), $\eta_0 \in \widetilde{CIS}_\eta(\boldsymbol{\phi}_0|\mathbf{s}(\mathbf{y}^T), N)$ holds for any $\mathbf{y}^T$. Hence, for all T,

$$
P_{\mathbf{Y}^T|\mathbf{s},\boldsymbol{\phi}}(\eta_0 \in \widehat{C}_\alpha^*|\mathbf{s}(\mathbf{Y}^T), \boldsymbol{\phi}_0) \geq P_{\mathbf{Y}^T|\boldsymbol{\phi}}(\widetilde{CIS}_\eta(\boldsymbol{\phi}_0|\mathbf{s}(\mathbf{Y}^T), N) \subset \widehat{C}_\alpha^*|\mathbf{s}(\mathbf{Y}^T), \boldsymbol{\phi}_0). \tag{D.2.1}
$$

Hence, to prove the claim, it suffices to focus on the asymptotic behavior of the coverage probability for the conditional identified set shown in the right-hand side.

Under Assumption 6.6.2 and 6.6.3, the asymptotically correct coverage for the conditional identified set can be obtained by applying Proposition 2 in GK. $\square$

Primitive Conditions for Assumption 6.6.3. In what follows, we present sufficient conditions for convexity, continuity and differentiability (both in $\boldsymbol{\phi}$) of the conditional impulse-response identified set under the assumption that there is a fixed number of shock-sign restrictions constraining the first structural shock only (possibly in multiple periods).

Proposition D.2.1. ***Convexity.*** *Let the parameter of interest be $\eta_{i,1,h}$, the impulse response of the ith variable at the hth horizon to the first structural shock. Assume that there are shock-sign restrictions on $\varepsilon_{1,t}$ for $t = t_1, \ldots, t_K$, so $N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{Y}^T) = (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1}, \ldots, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K})'\mathbf{q}_1 \geq \mathbf{0}_{K\times 1}$. Then the set of values of $\eta_{i,1,h}$ satisfying the shock-sign restrictions and sign normalization, $\{\eta_{i,1,h}(\boldsymbol{\phi}, \mathbf{Q}) = \mathbf{c}_{i,h}(\boldsymbol{\phi})\mathbf{q}_1 : N(\boldsymbol{\phi}, \mathbf{Q}, \mathbf{Y}^T) \geq \mathbf{0}_{K\times 1}, \mathrm{diag}(\mathbf{Q}'\boldsymbol{\Sigma}_{tr}^{-1}) \geq \mathbf{0}_{n\times 1}, \mathbf{Q} \in \mathcal{O}(n)\}$ is convex for all i and h if there exists a unit-length vector $\mathbf{q} \in \mathbb{R}^n$ satisfying*

$$\begin{bmatrix} (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1}, \ldots, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K})' \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} \geq \mathbf{0}_{(K+1)\times 1}. \tag{D.2.2}$$

Proof of Proposition D.2.1. If there exists a unit-length vector $\mathbf{q}$ satisfying the inequality in (D.2.2), it must lie within the intersection of the K half-spaces defined by the inequalities $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_k})'\mathbf{q} \geq 0$, $k = 1, \ldots, K$, the half-space defined by the sign normalization, $(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})'\mathbf{q} \geq 0$, and the unit sphere in $\mathbb{R}^n$. The intersection of these $K+1$ half-spaces and the unit sphere is a path-connected set. Since $\eta_{i,1,h}(\boldsymbol{\phi}, \mathbf{Q})$ is a continuous function of $\mathbf{q}_1$, the set of values of $\eta_{i,1,h}$ satisfying the restrictions is an interval and is thus convex, because the set of a continuous function with a path-connected domain is always an interval. □

Proposition D.2.2. ***Continuity.*** *Let the parameter of interest and restrictions be as in Proposition D.2.1, and assume that the conditions in the proposition are satisfied. If there exists a unit-length vector $\mathbf{q} \in \mathbb{R}^n$ such that, at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$,*

$$\begin{bmatrix} (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1}, \ldots, \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K})' \\ (\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n})' \end{bmatrix} \mathbf{q} >> \mathbf{0}_{(K+1)\times 1}, \tag{D.2.3}$$

then $u(\boldsymbol{\phi}, \mathbf{Y}^T)$ and $l(\boldsymbol{\phi}, \mathbf{Y}^T)$ are continuous at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ for all i and h.[1]

Proof of Proposition D.2.2. $\mathbf{Y}^T$ enters the NR through the reduced-form VAR innovations, $\mathbf{u}_t$. After noting that the reduced-form VAR innovations are (implicitly)

[1]For a vector $\mathbf{x} = (x_1, \ldots, x_m)'$, $\mathbf{x} >> \mathbf{0}_{m\times 1}$ means that $x_i > 0$ for all $i = 1, \ldots, m$.

continuous in $\boldsymbol{\phi}$, continuity of $u(\boldsymbol{\phi},\mathbf{Y}^T)$ and $l(\boldsymbol{\phi},\mathbf{Y}^T)$ follows by the same logic as in the proof of Proposition B.2 in the supplemental material of Giacomini and Kitagawa (2021). We omit the detail for brevity. □

Proposition D.2.3. ***Differentiability.*** *Let the parameter of interest and restrictions be as in Proposition D.2.1, and assume that the conditions in the proposition are satisfied. Denote the unit sphere in $\mathbb{R}^n$ by $\mathscr{S}^{n-1}$. If, at $\boldsymbol{\phi}=\boldsymbol{\phi}_0$, the set of solutions to the optimization problem*

$$\max_{\mathbf{q}\in\mathscr{S}^{n-1}}\left(\min_{\mathbf{q}\in\mathscr{S}^{n-1}}\right)\mathbf{c}_{i,h}'(\boldsymbol{\phi})\mathbf{q} \tag{D.2.4}$$

$$\text{s.t.}\quad \left[(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1},\ldots,\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K}),\quad \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n}\right]'\mathbf{q}\geq\mathbf{0}_{(K+1)\times 1} \tag{D.2.5}$$

is singleton, the optimized value $u(\boldsymbol{\phi},\mathbf{Y}^T)$ ($l(\boldsymbol{\phi},\mathbf{Y}^T)$) is nonzero, and the number of binding inequality restrictions at the optimum is at most $n-1$, then $u(\boldsymbol{\phi},\mathbf{Y}^T)$ ($l(\boldsymbol{\phi},\mathbf{Y}^T)$) is almost-surely differentiable at $\boldsymbol{\phi}=\boldsymbol{\phi}_0$.

Proof of Proposition D.2.3. One-to-one differentiable reparameterization of the optimization problem in Equation (D.2.4) using $\mathbf{x}=\boldsymbol{\Sigma}_{tr}\mathbf{q}$ yields the optimization problem in Equation (2.5) of Gafarov, Meier and Montiel Olea (2018) with a set of inequality restrictions that are now a function of the data through the reduced-form VAR innovations entering the NR. Noting that $\mathbf{u}_t$ is (implicitly) differentiable in $\boldsymbol{\phi}$, differentiability of $u(\boldsymbol{\phi},\mathbf{Y}^T)$ at $\boldsymbol{\phi}=\boldsymbol{\phi}_0$ follows from their Theorem 2 under the assumptions that, at $\boldsymbol{\phi}=\boldsymbol{\phi}_0$, the set of solutions to the optimization problem is singleton, the optimized value $u(\boldsymbol{\phi},\mathbf{Y}^T)$ is nonzero, and the number of binding sign restrictions at the optimum is at most $n-1$. Differentiability of $l(\boldsymbol{\phi},\mathbf{Y}^T)$ follows similarly. Note that Theorem 2 of Gafarov et al. (2018) additionally requires that the column vectors of $\left[(\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_1},\ldots,\boldsymbol{\Sigma}_{tr}^{-1}\mathbf{u}_{t_K}),\ \ \boldsymbol{\Sigma}_{tr}^{-1}\mathbf{e}_{1,n}\right]$ are linearly independent, but this occurs almost-surely under the probability law for $\mathbf{Y}^T$. □

Bibliography

ADAM, K. AND H. WEBER (2019): "Optimal Trend Inflation," *American Economic Review*, 109, 702–37.

ADJEMIAN, S., H. BASTANI, M. JUILLARD, F. KARAMÉ, J. MAIH, F. MIHOUBI, W. MUTSCHLER, G. PERENDIA, J. PFEIFER, M. RATTO, AND S. VILLEMOT (2020): "Dynare: Reference Manual Version 4," Dynare Working Papers 1, CEPREMAP.

AMIR-AHMADI, P. AND T. DRAUTZBURG (2021): "Identification and Inference with Ranking Restrictions," *Quantitative Economics*, 12, 1–39.

ANDREWS, I., M. GENTZKOW, AND J. M. SHAPIRO (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132, 1553–1592.

——— (2020): "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 88, 2231–2258.

ANGELINI, G. AND L. FANELLI (2019): "Exogenous Uncertainty and the Identification of Structural Vector Autoregressions with External Instruments," *Journal of Applied Econometrics*, 34, 951–971.

ANTOLÍN-DÍAZ, J. AND J. F. RUBIO-RAMÍREZ (2018): "Narrative Sign Restrictions for SVARs," *American Economic Review*, 108, 2802–29.

ARIAS, J. E., D. CALDARA, AND J. F. RUBIO-RAMÍREZ (2019): "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure," *Journal of Monetary Economics*, 101, 1–13.

ARIAS, J. E., J. F. RUBIO-RAMÍREZ, AND D. F. WAGGONER (2018): "Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications," *Econometrica*, 86, 685720.

——— (2021): "Inference in Bayesian Proxy-SVARs," *Journal of Econometrics*, 225, 88–106.

ARMSTRONG, T. AND M. KOLESÁR (2021): "Sensitivity Analysis using Approximate Moment Condition Models," *Quantitative Economics*, 12, 77–108.

BACCHIOCCHI, E. AND T. KITAGAWA (2020): "Locally- But Not Globally-Identified SVARs," cemmap Working Paper CWP40/20.

——— (2021): "A Note on Global Identification in Structural Vector Autoregressions," arXiv: 2102.04048 [econ.EM].

BAHAJ, S. (2020): "Sovereign Spreads in the Euro Area: Cross Border Transmission and Macroeconomic Implications," *Journal of Monetary Economics*, 110, 116–135.

BARANKIN, E. W. (1960): "Sufficient Parameters: Solution of the Minimal Dimensionality Problem," *Annals of the Institute of Statistical Mathematics*, 12, 91–118.

BASU, A., H. SHIOYA, AND C. PARK (2011): *Statistical Inference: The Minimum Distance Approach*, Chapman and Hall/CRC Press.

BASU, S. AND J. G. FERNALD (1997): "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 105, 249–283.

BAUMEISTER, C. AND J. D. HAMILTON (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963–1999.

——— (2018): "Inference in Structural Vector Autoregressions when the Identifying Assumptions are not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations," *Journal of Monetary Economics*, 100, 48 – 65.

——— (2019): "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review*, 109, 1873–1910.

BAYER, C., B. BORN, AND R. LUETTICKE (2019): "Shocks, Frictions, and Inequality in US Business Cycles," .

BEN ZEEV, N. (2018): "What Can We Learn About News Shocks from the Late 1990s and Early 2000s Boom-bust Period?" *Journal of Economic Dynamics and Control*, 87, 94–105.

BERGER, J. AND L. M. BERLINER (1986): "Robust Bayes and Empirical Bayes Analysis with ε-contaminated Priors," *The Annals of Statistics*, 14, 461–486.

BERGER, J. O. (1984): "The Robust Bayesian Viewpoint," in *Robustness of Bayesian Analysis*, ed. by J. Kadane, Amsterdam, Netherlands: North-Holland.

——— (1985): *Statistical Decision Theory and Bayesian Analysis*, New York, NY: Springer-Verlag, 2nd ed.

——— (1994): "An Overview of Robust Bayesian Analysis," *TEST*, 3, 5–58.

BERGIN, P. R. AND G. CORSETTI (2008): "The Extensive Margin and Monetary Policy," *Journal of Monetary Economics*, 55, 1222–1237.

BETRÒ, B. AND F. RUGGERI (1992): "Conditional Γ-minimax Actions Under Convex Losses," *Communications in Statistics, Part A – Theory and Methods*, 21, 1051–1066.

BILBIIE, F. O., I. FUJIWARA, AND F. GHIRONI (2014): "Optimal Monetary Policy with Endogenous Entry and Product Variety," *Journal of Monetary Economics*, 64, 1–20.

BILBIIE, F. O., F. GHIRONI, AND M. J. MELITZ (2007): "Monetary Policy and Business Cycles with Endogenous Entry and Product Variety," *NBER Macroeconomics Annual*, 22, 299–353.

BLANCHARD, O. J. AND D. QUAH (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79, 655–673.

BONHOMME, S. AND M. WEIDNER (2018): "Minimizing Sensitivity to Model Misspecification," cemmap Working Paper CWP59/18.

BOYD, S. AND L. VANDENBERGHE (2004): *Convex Optimization*, Cambridge: Cambridge University Press.

BRAUN, R. AND R. BRÜGGEMANN (2017): "Identication of SVAR Models by Combining Sign Restrictions With External Instruments," University of Konstanz Department of Economics Working Paper Series 2017-07.

BROWN, L. D. AND R. PURVES (1973): "Measurable Selections of Extrema," *Annals of Statistics*, 1, 902–912.

CALDARA, D. AND E. HERBST (2019): "Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs," *American Economic Journal: Macroeconomics*, 11, 157–92.

CALDARA, D. AND C. KAMPS (2017): "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers," *The Review of Economic Studies*, 84, 1015–1040.

CHAMBERLAIN, G. (2000): "Econometric Applications of Maxmin Expected Utilities," *Journal of Applied Econometrics*, 15, 625–644.

CHAMBERLAIN, G. AND E. E. LEAMER (1976): "Matrix Weighted Averages and Posterior Bounds,," *Journal of the Royal Statistical Society. Series B (Methodological)*, 38, 73–84.

CHEN, X., T. CHRISTENSEN, AND E. TAMER (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86, 1965 – 2018.

CHENG, K. AND Y. YANG (2020): "Revisiting the Effects of Monetary Policy Shocks: Evidence from SVAR with Narrative Sign Restrictions," *Economics Letters*, 196, 109598.

CHIB, S. AND I. JELIAZKOV (2001): "Likelihoods from the Metropolis Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

CHRISTENSEN, T. AND B. CONNAULT (2021): "Counterfactual Sensitivity and Robustness," arXiv:1904.00989 [econ.EM].

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (1999): "Chapter 2 – Monetary Policy Shocks: What Have We Learned and to What End?" in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, Elsevier, vol. 1, 65–148.

CLEMENTI, G. L. AND B. PALAZZO (2016): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journal: Macroeconomics*, 8, 1–41.

DASGUPTA, A. (2008): *Asymptotic Theory of Statistics and Probability*, Springer Texts in Statistics, Springer-Verlag New York.

DASGUPTA, A. AND W. J. STUDDEN (1989): "Frequentist Behavior of Robust Bayes Estimates of Normal Means," *Statistics and Decisions*, 7, 333–361.

DEL NEGRO, M. AND F. SCHORFHEIDE (2011): "Bayesian Macroeconometrics," in *Oxford Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. V. Dijk, Oxford University Press, 293389.

DENNEBERG, D. (1994): *Non-additive Measure and Integral*, Dordrecht, Netherlands: Kluwer Academic Publishers.

DEY, D. K. AND A. C. MICHEAS (2000): "Ranges of Posterior Expected Losses and ε-Robust Actions," in *Robust Bayesian Analysis*, ed. by D. Ríos Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 145–159.

DRAUTZBURG, T. (2020): "A Narrative Approach to a Fiscal DSGE Model," *Quantitative Economics*, 11, 801–837.

ELLSBERG, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *The Quarterly Journal of Economics*, 75, 643–669.

FAVARA, G., S. GILCHRIST, K. F. LEWIS, AND E. ZAKRAJSĚK (2016): "Updating the Recession Risk and the Excess Bond Premium," Board of governors of the federal reserve system feds notes.

FERNALD, J. G. (2012): "A Quarterly, Utilization-adjusted Series on Total Factor Productivity," Federal Reserve Bank of San Francisco Working Paper Series 2012-19.

FOSTER, L., J. C. HALTIWANGER, AND C. J. KRIZAN (2001): "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in *New Developments in Productivity Analysis*, ed. by C. R. Hulten, E. R. Dean, and M. J. Harper, University of Chicago Press, 303–372.

FURLANETTO, F. AND Ø. ROBSTAD (2019): "Immigration and the Macroeconomy: Some New Empirical Evidence," *Review of Economic Dynamics*, 34, 1–19.

GAFAROV, B., M. MEIER, AND J. L. MONTIEL OLEA (2018): "Delta-Method Inference for a Class of Set-Identified SVARs," *Journal of Econometrics*, 203, 316–327.

GALÍ, J. (2008): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press.

GERTLER, M. AND P. KARADI (2015): "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

GEWEKE, J. (1999): "Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18, 1–126.

GIACOMINI, R. AND T. KITAGAWA (2021): "Robust Bayesian Inference for Set-identified Models," *Econometrica*, 29, 1519–1556.

GIACOMINI, R., T. KITAGAWA, AND M. READ (2021): "Identification and Inference Under Narrative Restrictions," arXiv: 2102.06456 [econ.EM].

——— (2022a): "Robust Bayesian Inference in Proxy SVARs," *Journal of Econometrics*, 228, 107–126.

GIACOMINI, R., T. KITAGAWA, AND H. UHLIG (2019): "Estimation Under Ambiguity," cemmap Working Paper CWP24/19.

GIACOMINI, R., T. KITAGAWA, AND A. VOLPICELLA (2022b): "Uncertain Identification," *Quantitative Economics*, 13, 95–123.

GILBOA, I. AND D. SCHMEIDLER (1993): "Updating Ambiguous Beliefs," *Journal of Economic Theory*, 59, 33–49.

GILCHRIST, S. AND E. ZAKRAJŠEK (2012): "Credit Spreads and Business Cycle Fluctuations," *The American Economic Review*, 102, 1692–1720.

GOOD, I. J. (1965): *The Estimation of Probabilities*, Cambridge, MA: MIT Press.

GRANZIERA, E., H. R. MOON, AND F. SCHORFHEIDE (2018): "Inference for VARs Identified with Sign Restrictions," *Quantitative Economics*, 9, 1087–1121.

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

GUSTAFSON, P. (2000): "Local Robustness in Bayesian Analysis," in *Robust Bayesian Analysis*, ed. by D. Ríos Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 71–88.

HAMILTON, J. D. (1994): *Time Series Analysis*, Princeton University Press.

HANSEN, L. P. AND T. J. SARGENT (2001): "Robust Control and Model Uncertainty," *American Economic Review, AEA Papers and Proceedings*, 91, 60–66.

HO, P. (2020): "Global Robust Bayesian Analysis in Large Models," Federal Reserve Bank of Richmond WP 20-07.

HOPENHAYN, H. AND R. ROGERSON (1993): "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, 101, 915–938.

HOPENHAYN, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60, 1127–1150.

HUBER, P. J. (1973): "The Use of Choquet Capacities in Statistics," *Bulletin of the International Statistical Institute*, 45, 181–191.

HUBER, P. J. AND E. M. RONCHETTI (2009): *Robust Statistics, 2nd edition*, John Wiley and Sons.

INOUE, A. AND L. KILIAN (2020): "Joint Bayesian Inference about Impulse Responses in VAR Models," .

JAROCIŃSKI, M. AND P. KARADI (2020): "Deconstructing Monetary Policy Surprises – The Role of Information Shocks," *American Economic Journal: Macroeconomics*, 12, 1–43.

JENTSCH, C. AND K. G. LUNSFORD (2019): "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment," *American Economic Review*, 109, 2655–2678.

KHAN, A. AND J. K. THOMAS (2008): "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics," *Econometrica*, 76, 395–436.

KILIAN, L. AND H. LÜTKEPOHL (2017): *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.

KILIAN, L. AND D. P. MURPHY (2012): "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models," *Journal of the European Economic Association*, 10, 1166–1188.

KILIAN, L. AND X. ZHOU (2020a): "Does Drawing Down the US Strategic Petroleum Reserve Help Stabilize Oil Prices?" *Journal of Applied Econometrics*, 35, 673–691.

——— (2020b): "Oil Prices, Exchange Rates and Interest Rates," Center for financial studies working paper series no. 646.

KITAMURA, Y., T. OTSU, AND K. EVDOKIMOV (2013): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 81, 1185–1201.

KLINE, B. AND E. TAMER (2016): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7, 329–366.

KOMAROVA, T. (2013): "Binary Choice Models with Discrete Regressors: Identification and Misspecification," *Journal of Econometrics*, 177, 14–33.

KOOPMANS, T. C. AND R. REIERSOL (1950): "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21, 165–181.

KOPECKY, K. A. AND R. M. H. SUEN (2010): "Finite State Markov-chain Approximations to Highly Persistent Processes," *Review of Economic Dynamics*, 13, 701 – 714.

KOSOROK, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*, Springer Series in Statistics, Springer-Verlag New York.

KRUSELL, P. AND A. A. SMITH (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867–896.

KUDŌ, H. (1967): "On Partial Prior Information and the Property of Parametric Sufficiency," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by L. Le Cam and J. Neyman, vol. 1, 251–265.

LAUMER, S. (2020): "Government Spending and Heterogeneous Consumption Dynamics," *Journal of Economic Dynamics and Control*, 114, 103868.

LAVINE, M., L. WASSERMAN, AND R. L. WOLPERT (1991): "Bayesian Inference with Specified Prior Marginals," *Journal of the American Statistical Association*, 86, 964–971.

LEAMER, E. E. (1978): *Specification Searches*, New York, NY: Wiley.

——— (1982): "Sets of Posterior Means with Bounded Variance Priors," *Econometrica*, 50, 725–736.

LEE, Y. (2005): "The Importance of Reallocations in Cyclical Productivity and Returns to Scale: Evidence from Plant-Level Data," Federal Reserve Bank of Cleveland Working Paper 0509.

LEE, Y. AND T. MUKOYAMA (2018): "A Model of Entry, Exit, and Plant-Level Dynamics Over the Business Cycle," *Journal of Economic Dynamics and Control*, 96, 1–25.

LEWIS, V. (2006): "Macroeconomic Fluctuations and Firm Entry: Theory and Evidence," National Bank of Belgium Working Paper No. 103.

LEWIS, V. AND C. POILLY (2012): "Firm Entry, Markups and the Monetary Transmission Mechanism," *Journal of Monetary Economics*, 59, 670–685.

LI, Y. AND S. K. GHOSH (2015): "Efficient Sampling Methods for Truncated Multivariate Normal and Student-t Distributions Subject to Linear Inequality Constraints," *Journal of Statistical Theory and Practice*, 9, 712–732.

LIAO, Y. AND A. SIMONI (2013): "Semi-parametric Bayesian Partially Identified Models Based on Support Function," arXiv:1212.3267 [stat.ME].

LUDVIGSON, S. C., S. MA, AND S. NG (2018): "Shock Restricted Structural Vector-Autoregressions," National Bureau of Economic Research Working Paper No. 23225.

——— (2021): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics*, 13, 369–410.

LUNSFORD, K. G. (2015): “Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy,” Federal Reserve Bank of Cleveland Working Paper 15-28.

MANGASARIAN, O. L. (1969): *Nonlinear Programming*, NY: McGraw-Hill.

MANSKI, C. F. (1975): “Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 3, 205–228.

——— (1981): “Learning and Decision Making when Subjective Probabilities have Subjective Domains,” *Annals of Statistics*, 9, 59–65.

——— (1985): “Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator,” *Journal of Econometrics*, 27, 313–333.

——— (2000): “Identification Problems and Decisions Under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice,” *Journal of Econometrics*, 95, 415–442.

MERTENS, K. AND J. L. MONTIEL OLEA (2018): “Marginal Tax Rates and Income: New Time Series Evidence,” *The Quarterly Journal of Economics*, 133, 1803–1884.

MERTENS, K. AND M. O. RAVN (2013): “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States,” *American Economic Review*, 103, 1212–47.

——— (2014): “A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers,” *Journal of Monetary Economics*, 68, S1–S19.

——— (2019): “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Reply,” *American Economic Review*, 109, 2679–91.

MOLCHANOV, I. (2005): *Theory of Random Sets*, London, United Kingdom: Springer-Verlag.

MOLCHANOV, I. AND F. MOLINARI (2018): *Random Sets in Econometrics*, Econometric Society Monographs, Cambridge, United Kingdom: Cambridge University Press.

MONTIEL OLEA, J. L., J. H. STOCK, AND M. W. WATSON (2021): "Inference in Structural Vector Autoregressions Identified with an External Instrument," *Journal of Econometrics*, 225, 74–87.

MOON, H. R. AND F. SCHORFHEIDE (2011): "Bayesian and Frequentist Inference in Partially Identified Models," NBER Working Paper No 14882.

——— (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755–782.

MORAN, P. AND A. QUERALTO (2018): "Innovation, Productivity, and Monetary Policy," *Journal of Monetary Economics*, 93, 24–41.

MORENO, E. AND J. A. CANO (1995): "Classes of Bidimensional Priors Specified on a Collection of Set: Bayesian Robustness," *Journal of Statistical Inference and Planning*, 46, 325–334.

MÜLLER, U. K. (2012): "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models," *Journal of Monetary Economics*, 59, 581–597.

NORETS, A. AND X. TANG (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *Review of Economic Studies*, 81, 1229–1262.

OTTONELLO, P. AND T. WINBERRY (2020): "Financial Heterogeneity and the Investment Channel of Monetary Policy," *Econometrica*, 88, 2473–2502.

PETERSON, I. R., M. R. JAMES, AND P. DUPUIS (2000): "Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints," *ISSS Transactions on Automatic Control*, 45, 398–412.

PETTERSON, M. S., D. SEIM, AND J. M. SHAPIRO (2020): "Bounds on a Slope from Size Restrictions on Economic Shocks," .

PIFFER, M. AND M. PODSTAWSKI (2018): "Identifying Uncertainty Shocks Using the Price of Gold," *The Economic Journal*, 128, 3266–3284.

PIRES, C. P. (2002): "A Rule for Updating Ambiguous Beliefs," *Theory and Decision*, 33, 137–152.

PLAGBORG-MØLLER, M. (2019): "Bayesian Inference on Structural Impulse Response Functions," *Quantitative Economics*, 10, 145–184.

PLAGBORG-MØLLER, M. AND C. WOLF (in press): "Instrumental Variable Identification of Dynamic Variance Decompositions," *Journal of Political Economy*.

PLAGBORG-MØLLER, M. AND C. K. WOLF (2021a): "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 89, 955–980.

——— (2021b): "Supplement to "Local Projections and VARs Estimate the Same Impulse Responses"," *Econometrica*.

POIRIER, D. J. (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14, 483–509.

PUGSLEY, B. W., P. SEDLÁČEK, AND V. STERK (2021): "The Nature of Firm Growth," *American Economic Review*, 111, 547–579.

RAMEY, V. A. (2016): "Chapter 2 – Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 71 – 162.

READ, M. (in press): "Algorithms for Inference in SVARs Identified with Sign and Zero Restrictions," *Econometrics Journal*.

REDL, C. (2020): "Uncertainty Matters: Evidence from Close Elections," *Journal of International Economics*, 103296.

RIEDER, H. (1994): *Robust Asymptotic Statistics*, New York, NY: Springer.

RÍOS INSUA, D. AND F. RUGGERI, eds. (2000): *Robust Bayesian Analysis*, New York, NY: Springer.

ROBBINS, H. (1956): "An Empirical Bayes Approach to Statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 157–163.

ROMER, C. D. AND D. H. ROMER (1989): "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," in *NBER Macroeconomics Annual*, ed. by O. J. Blanchard and S. Fischer, Cambridge, MA: MIT Press, vol. 4, 121–84.

——— (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.

——— (2010): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100, 763–801.

ROSEN, A. M. AND T. URA (2020): "Finite Sample Inference for the Maximum Score Estimand," cemmap Working Paper CWP22/20.

ROTEMBERG, J. (1982): "Sticky Prices in the United States," *Journal of Political Economy*, 90, 1187–1211.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

ROUWENHORST, K. G. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Frontiers of Business Cycle Research*, ed. by T. Cooley, Princeton University Press, Princeton N.J., 294 – 330.

RUBIO-RAMÍREZ, J. F., D. F. WAGGONER, AND T. ZHA (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *The Review of Economic Studies*, 77, 665–696.

SADEGHI, A. (2008): "The Births and Deaths of Business Establishments in the United States," *Monthly Labor Review*, December 2008, 3–18.

SAMANIEGO, R. M. (2008): "Entry, Exit and Business Cycles in a General Equilibrium Model," *Review of Economic Dynamics*, 11, 529–541.

SCHERVISH, M. J. (1995): *Theory of Statistics*, Springer Series in Statistics, Springer-Verlag New York.

SCHMEIDLER, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57, 571–587.

SCHMITT-GROHÉ, S. AND M. URIBE (2014): "Solving Dynamic General Equilibrium Models Using a Second-order Approximation to the Policy Function," *Journal of Economic Dynamics and Control*, 28, 755–775.

SHYAMALKUMAR, N. D. (2000): "Likelihood Robustness," in *Robust Bayesian Analysis*, ed. by D. Ríos Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 127–143.

SIMS, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

SIMS, C. A., D. F. WAGGONER, AND T. ZHA (2008): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*, 146, 255–274.

SIVAGANESAN, S. AND J. O. BERGER (1989): "Ranges of Posterior Measures for Priors with Unimodal Contaminations," *Annals of Statistics*, 17, 868–889.

STAIGER, D. AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H. (2008): "What's New in Econometrics–Time Series, Lecture 7: Structural VARs," Cambridge, MA: National Institute for Economic Research.

STOCK, J. H. AND M. W. WATSON (2012): "Disentangling the Channels of the 2007–09 Recession," *Brookings Papers on Economic Activity,*, Spring, 81–156.

——— (2016): "Chapter 8 – Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in

Handbook of Macroeconomics, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 415 – 525.

——— (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 128, 917–948.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20, 518–529.

THOMAS, J. K. (2002): "Is Lumpy Investment Relevant for the Business Cycle?" *Journal of Political Economy*, 110, 508–534.

UHLIG, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

——— (2017): "Shocks, Sign Restrictions, and Identification," in *Advances in Economics and Econometrics: Eleventh World Congress*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press, vol. 2, 95127.

UUSKÜLA, L. (2016): "Monetary Transmission Mechanism with Firm Turnover," *Journal of Macroeconomics*, 50, 1–18.

VIDAKOVIC, B. (2000): "Γ-Minimax: A Paradigm for Conservative Robust Bayesians," in *Robust Bayesian Analysis*, ed. by D. Ríos Insua and F. Ruggeri, Springer, Lecture Notes in Statistics, 241–259.

WALD, A. (1950): *Statistical Decision Functions*, New York, NY: Wiley.

WASSERMAN, L. A. (1990): "Prior Envelopes Based on Belief Functions," *The Annals of Statistics*, 18, 454–464.

WATSON, J. AND C. HOLMES (2016): "Approximate Models and Robust Decisions," *Statistical Science*, 31, 465–489.

ZHOU, X. (2020): "Refining the Workhorse Oil Market Model," *Journal of Applied Econometrics*, 35, 130–140.

Statement of Conjoint Work

Note on the joint work in Matthew Read's thesis "Essays on the Effects of Macroeconomic Shocks".

Chapter 2, "Monetary Policy and Firm Dynamics", is single-authored by Matthew Read.
Chapter 3, "Algorithms for Inference in SVARs Identified with Sign and Zero Restrictions", is single-authored by Matthew Read.
Chapter 4, "Robust Bayesian Analysis for Econometrics", was undertaken as joint work with Raffaella Giacomini and Toru Kitagawa.
Chapter 5, "Robust Bayesian Inference in Proxy SVARs", was undertaken as joint work with Raffaella Giacomini and Toru Kitagawa.
Chapter 6, "Identification and Inference Under Narrative Restrictions", was undertaken as joint work with Raffaella Giacomini and Toru Kitagawa.

Contributions from the authors were equal in the case of the coauthored chapters.
Signatures from coauthors: