

SVARs With Occasionally-Binding Constraints *

S. Boragan Aruoba
University of Maryland

Marko Mlikota
University of Pennsylvania

Frank Schorfheide
University of Pennsylvania
CEPR, NBER, PIER

Sergio Villalvazo
University of Pennsylvania

Current Version: April 11, 2021

Abstract

We develop a structural VAR in which an occasionally-binding constraint generates censoring of one of the dependent variables. Once the censoring mechanism is triggered, we allow some of the coefficients for the remaining variables to change. We show that a necessary condition for a unique reduced form is that regression functions for the non-censored variables are continuous at the censoring point and that parameters satisfy some mild restrictions. In our application the censored variable is a nominal interest rate constrained by an effective lower bound (ELB). According to our estimates based on U.S. data, once the ELB becomes binding, the coefficients in the inflation equation change significantly, which translates into a change of the inflation responses to (unconventional) monetary policy and demand shocks. Our results suggest that the presence of the ELB is indeed empirically relevant for the propagation of shocks. We also obtain a shadow interest rate that shows a significant accommodation in the early parts of the Great Recession, followed by a mild and steady accommodation until liftoff in 2016.

JEL CLASSIFICATION: C11, C22, C34, E32, E52

KEY WORDS: Bayesian inference, effective lower bound, limited dependent variables, sequential Monte Carlo methods, structural VARs, shadow rate.

* Correspondence: B. Aruoba (aruoba@umd.edu), Department of Economics, University of Maryland, College Park, MD 20742. M. Mlikota (mlikota@sas.upenn.edu), F. Schorfheide (schorf@ssc.upenn.edu), S. Villalvazo (vsergio@sas.upenn.edu): Department of Economics, University of Pennsylvania, 133 S. 36th Street, Philadelphia, PA 19104. We are thankful for helpful comments and suggestions from the guest editor, two referees, Sophocles Mavroeidis, Mathias Trabandt, and seminar participants at the Federal Reserve Board, Maryland, and the Schumpeter Seminar (Berlin). Aruoba and Schorfheide gratefully acknowledge financial support from the National Science Foundation under Grant SES 1851634.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are widely used in central banks, by regulators, and in academia to study the effects of monetary and macroprudential policies and the propagation of shocks in the macro economy. The most recent vintage of these models involves occasionally-binding constraints arising from, for instance, an effective lower bound (ELB) on nominal interest rates. Agents' decision rules in these models typically exhibit "kinks," meaning the elasticity of the choice variables, say output or prices, with respect to the underlying state variables changes drastically when the constraint becomes binding. We used this observation in Aruoba, Cuba-Borda, Higa-Flores, Schorfheide, and Villalvazo (2021), henceforth ACHSV, to approximate such decision rules through piecewise-linear and continuous (PLC) functions. The model solution has the form of a vector autoregression (VAR) with a censored dependent variable and regime switching coefficients. The regime shift is endogenous and it is linked to the censoring mechanism. For instance, in a model with an ELB constraint, the VAR coefficients switch once the interest rate reaches the lower bound.

Our paper develops a structural VAR (SVAR) that mimics the PLC-DSGE model solution, but can be used independently of an optimization-based structural model to study the propagation of shocks in settings where an observable is subject to an occasionally-binding constraint. Throughout this paper, we focus on an application that features nominal interest rates that are constrained by an ELB.¹ An important empirical question is whether the propagation of shocks works differently when the economy reaches the ELB because interest rates no longer can fall in response to adverse shocks and agents may adjust their behavior in light of the constraint. In densely parameterized models such as time-varying coefficient structural VARs (TVC-SVARs), these effects are empirically difficult to measure because for many countries we only have a few years of ELB observations available. This makes it challenging to obtain precise estimates. Our proposed model is able to avoid this problem because the coefficient shift that takes place once the constrained becomes binding is controlled by a low-dimensional vector of additional parameters.

¹We use the concept of effective instead of zero lower bound (ZLB). In a DSGE model that explicitly models money demand, such as the one in Aruoba and Schorfheide (2011), an interest rate less than zero means a monetary equilibrium ceases to exist. One can generalize these models to include storage cost of physical money and allow for the interest rate to go below zero. In fact the Bank of Japan and the European Central Bank, among other central banks, have been able to reduce their policy rates below zero. Nonetheless, it remains plausible to assume that there is a bound beyond which it becomes very difficult to lower interest rates further and this is what the literature considers to be the ELB.

There is a small, but growing literature on multivariate time series models with a censored interest rate variable, including Iwata and Wu (2006), Ikeda, Li, Mavroeidis, and Zanetti (2020), Mavroeidis (2020), Carriero, Clark, Marcellino, and Mertens (2021) and Johansen and Mertens (2021). All of these models distinguish between a shadow rate, $y_{1,t}^*$ in our notation, and the actual interest rate $y_{1,t} = \max\{y_{1,t}^*, c\}$. The models differ, however, in terms of assumptions about the extent to which model coefficients are allowed to change when the economy reaches the ELB.

The key departure of our paper from the existing literature is the assumption that the coefficients that characterize the behavior of the private-sector variables $y_{2,t}$ (output gap and inflation in our application) as a function of lagged dependent variables and current structural shocks are allowed to switch once the ELB becomes binding. We define the endogenous regime indicator variable $s_t = \mathbb{I}\{y_{1,t} > c\}$. The regime-dependency of the coefficients is akin to capturing nonlinearities in decision rules that arise in DSGE model with occasionally-binding constraints.

The specification of a dynamic multivariate model with censoring and regime-dependent coefficients faces two challenges: parsimony and the existence of a unique reduced form. We show in this paper that these challenges are closely connected. The existence of a unique reduced form is referred to in the literature on censored simultaneous equations models as coherency and completeness; see Mavroeidis (2020). In every period t we can compute two hypothetical values for $y_t = [y_{1,t}, y'_{2,t}]'$: one is based on the $s_t = 1$ coefficients and the other one is based on the $s_t = 0$ coefficients. For y_t to be uniquely determined, it is necessary that only one of these two values has the property that the indicator function $\mathbb{I}\{y_{1,t} > c\}$ coincides with the s_t that was used to compute y_t . Typically, this uniqueness cannot be achieved without restricting the domain of the innovations ϵ_t .²

We prove three theoretical results. (i) To obtain uniqueness without domain restrictions for ϵ_t , it is necessary that the reduced-form regression functions are continuous whenever s_t switches between 0 and 1. The continuity requirement automatically imposes parsimony. We show that in a VAR with an n -dimensional vector y_t , the restriction reduces the number of free coefficients for the second regime from $(n-1) \times k$, where k is the number of regressors in each private-sector equation to $n-1$. (ii) Continuity is not sufficient for uniqueness without further coefficient restrictions. We provide a set of parameter restrictions for our SVAR specification that are necessary and sufficient to obtain uniqueness for all ϵ_t . (iii)

²In a VAR setting, the domain restriction would also be dependent on the lagged values of y_t .

We show that once continuity is imposed, the private sector equations can be rewritten as a simultaneous relationship between $y_{1,t}$, $y_{2,t}$, and the shadow rate $y_{1,t}^*$ with constant coefficients on current and lagged variables. Thus, in our application the piecewise-linear private-sector regression functions can be re-interpreted as agents reacting to a linear combination of the actual interest rate $y_{1,t}$ and the shadow rate $y_{1,t}^*$. Results (i) and (iii) are new. Conditional on having established (iii), it can be shown that (ii) reproduces a result in Mavroeidis (2020).

There is a debate in the literature whether lags of the censored interest rate y_{1t} or the shadow rate y_{1t}^* should appear in the conditional mean function. While this does not matter for the aforementioned theoretical results, it affects the estimation and identification of the model. In the DSGE model literature, authors often use lagged actual interest rates in the specification of the monetary policy rule, e.g., ACHSV, which means that the lagged actual interest rate becomes a state variable for private-sector decisions. This is the approach we follow in the empirical analysis.

Our empirical model also allows for stochastic volatility in the structural shocks. We use a bootstrap particle filter (BSPF) to integrate out the latent volatility from the likelihood function. We combine the likelihood function with a prior distribution and conduct Bayesian inference. Due to the presence of censoring, cross-regime coefficient restrictions, and stochastic volatility, the posterior distribution of the model parameters is non-standard. We use a sequential Monte Carlo (SMC) algorithm to draw from the posterior distribution; see Herbst and Schorfheide (2014, 2015) for DSGE applications and Bognanni and Herbst (2018) for an application to an SVAR with exogenous regime switches. Because the likelihood function inside the posterior sampler is evaluated using the BSPF, the resulting algorithm belongs to the SMC² family, studied by Chopin, Jacob, and Papaspiliopoulos (2013).

Abstracting from the nonlinearities generated by the censoring of the nominal interest rates, piecewise-linear regression equations for private-sector variables, and stochastic volatility, the specification of the SVAR follows the three-variable model estimated in Baumeister and Hamilton (2018), henceforth BH. While the prior distribution for the VAR coefficients is not identical to the one used by BH, the elicitation for the coefficients that describe the contemporaneous interaction between output, inflation, and interest rates follows a similar logic. The prior combines beliefs about aggregate demand and supply elasticities formed based on a simple New Keynesian DSGE model with beliefs about directions of impulse responses. As in BH, in the absence of observations from the ELB regime, our model is only set-identified. However, as shown in Mavroeidis (2020), the ELB regime generates additional identifying information and in our case leads to point identification.

We estimate our model based on quarterly U.S. data from 1984:Q1 to 2018:Q4. The vector y_t in our empirical analysis comprises the federal funds rate as the censored variable, output gap, and inflation. Based on the inspection of the posterior distribution and a marginal data density (MDD) comparison between a restricted and an unrestricted specification, we find evidence in favor of a kink in the inflation regression function. A comparison of impulse responses conditional on 1999:Q1 data (the U.S. was far away from the ELB) and conditional on 2009:Q1 data (the U.S. was at the ELB) yields the following findings: a negative monetary policy shock that creates the same size response in the shadow rate is more inflationary at the ELB on impact. This shock also generates a slightly larger response of the output gap, though credible intervals conditional on the two dates overlap. The response of inflation to negative demand shocks differ significantly when the ELB is non-binding and binding: Inflation falls in the former case but it slightly rises in the latter case. The difference between the responses persists for about three years. The output gap responses for demand and supply shocks do not significantly differ at and away from the ELB.

Our paper is related to several strands of the literature. From a methodological perspective, the paper most closely related to our work is Mavroeidis (2020). He also considers an SVAR with a censored dependent variable, or an occasionally-binding constraint, to capture the ELB constraint on nominal interest rates. He specifies the private sector equations as simultaneous relationship between the censored variable $y_{1,t}$, the latent variable $y_{1,t}^*$, and $y_{2,t}$. As discussed above, it turns out that our model with continuity imposed on the private-sector regression functions is identical to his specification, except in the following dimensions: our empirical specification allows for heteroskedastic structural shocks, which is important for inference on switches in the conditional mean function. Mavroeidis' specification allows for both lags of $y_{1,t}$ and the latent variable $y_{1,t}^*$ on the right-hand side of the VAR specification, whereas we stay close to the DSGE model specification in ACHSV and only use the former. A detailed comparison is provided in a separate subsection of our paper.

Building on an older literature on simultaneous equations models with censored dependent variables, e.g., Nelson and Olsen (1978), Gouriéroux, Laffont, and Monfort (1980), and Blundell and Smith (1989), the emphasis in Mavroeidis (2020) is on the identifying information that the censoring provides for the propagation of structural shocks. Important for the identification is whether, in the case of the ELB application, the private-sector variables respond to the actual censored interest rate or the uncensored shadow interest rate. In our

application, we let inflation and output gap respond to monetary policy shocks³, even if the economy is at the ELB, which is similar in spirit to the censored SVAR specification in Mavroeidis (2020) in which agents respond to a shadow rate. The identifying information encoded in the censoring mechanism is implicitly exploited in our Bayesian estimation through the updating of the SVAR coefficients based on the likelihood function.

Our empirical application focuses on the propagation of structural shocks in the U.S. at and away from the ELB. In this regard, it is closely related to Debortoli, Galí, and Gambetti (2019), Ikeda, Li, Mavroeidis, and Zanetti (2020), and Johansson and Mertens (2021). Debortoli, Galí, and Gambetti (2019) estimate an SVAR with exogenously time-varying coefficients (TVC-SVAR) using longer-term interest rates that do not reach the ELB. Based on their densely parameterized model, they do not find discernible differences between responses at and away from the ELB, which leads them to conclude that the ELB is empirically irrelevant. Our model allows for time-variation in coefficients in a much more parsimonious way and enables us to detect a significant change in coefficients that alters in particular the propagation of shocks to inflation. The analyses in Ikeda, Li, Mavroeidis, and Zanetti (2020) and Johansson and Mertens (2021) focus on the propagation of monetary policy shocks whereas we also consider supply and demand shocks. Our results on monetary policy shocks are qualitatively consistent with Johansson and Mertens (2021) who find that monetary accommodation during the recent ELB spell in the U.S. would have provided more stimulus than in other times.

The remainder of the paper is organized as follows. The specification of our SVAR with a censored dependent-variable and state-dependent regression functions is presented in Section 2. Our prior distribution for the SVAR parameters is discussed in Section 3. The likelihood function is derived in Section 4 and the SMC algorithm to implement the posterior computations is summarized in Section 5. Section 6 presents the empirical analysis and Section 7 concludes. Theoretical derivations and additional empirical results are relegated to the Online Appendix.

³This is also the case in DGSE model solutions; see ACHSV

2 SVAR Specification

We are using a structural vector autoregression (SVAR) to model the law of motion of the $n \times 1$ vector \tilde{y}_t . We express the SVAR in terms of deviations from a mean μ and define⁴

$$y_t = \tilde{y}_t - \mu.$$

There are two common ways of normalizing the coefficients of an SVAR: by setting the matrix of coefficients associated with the time t endogenous variables y_t equal to the identity matrix, or by setting the matrix that governs the impact of the structural shocks ϵ_t equal to the identity matrix. We begin with the former normalization and call it the Φ representation:

$$y_t = \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \Phi_c + \Phi_\epsilon \epsilon_t, \quad (1)$$

where $\epsilon_t \sim N(0, D)$ is a vector of structural innovations and D is a diagonal matrix. Even though in the current specification it is not separately identifiable from μ , we keep the intercept Φ_c in the notation. We will use it later when we introduce regime-dependent private-sector behavior. Reduced-form innovations can be defined as

$$u_t = \Phi_\epsilon \epsilon_t, \quad u_t \sim N(0, \Sigma) \text{ where } \Sigma = \Phi_\epsilon D \Phi_\epsilon'. \quad (2)$$

Define the $1 \times k$ vector $x_t' = [y_{t-1}', \dots, y_{t-p}', 1]$, $\Phi = [\Phi_1, \dots, \Phi_p, \Phi_c]'$, and $\Phi^\epsilon = \Phi_\epsilon'$ so that we can write the VAR as

$$y_t' = x_t' \Phi + \epsilon_t' \Phi^\epsilon. \quad (3)$$

Let $A = (\Phi^\epsilon)^{-1}$ and multiply (3) by A to re-normalize the VAR and obtain what we call the A representation of the VAR:

$$y_t' A = x_t' B + \epsilon_t', \quad (4)$$

where $B = \Phi A$.

Starting point for the subsequent analysis will be a third representation, which we refer to as $A\Phi$ representation. It combines a monetary policy rule written in A form with the private-sector equations in Φ form and has been used, for instance, in Del Negro and Schorfheide (2009). The Φ -form of the private-sector equations resemble the decision rules in a DSGE model solution that are allowed to differ at and away from the ELB. This representation

⁴The time invariant mean μ could also be replaced by a deterministic trend function μ_t .

also facilitates our proof of the existence of a unique reduced form; see Proposition 1 in Section 2.3 below.

Partition $y'_t = [y_{1,t}, y'_{2,t}]$ and $\epsilon'_t = [\epsilon_{1,t}, \epsilon'_{2,t}]$, where $y_{1,t}$ corresponds to the interest rate and $\epsilon_{1,t}$ is the monetary policy shock. Moreover, partition $B = [B_{\cdot 1}, B_{\cdot 2}]$, where $B_{\cdot 1}$ is a column vector that stacks the coefficients of the $y_{1,t}$ equation and the columns of the matrix $B_{\cdot 2}$ stack the coefficients for the private-sector equations. Finally, partition $A_{\cdot 1} = [A_{11} | A_{21}]$, where we use $|$ to indicate that the partitions are stacked. Using this notation, the monetary policy rule becomes

$$y_{1,t}A_{11} + y'_{2,t}A_{21} = x'_t B_{\cdot 1} + \epsilon_{1,t}. \quad (5)$$

Similarly, let $\Phi = [\Phi_{\cdot 1}, \Phi_{\cdot 2}]$ and $\Phi_{\cdot 2}^\epsilon = [\Phi_{12}^\epsilon, \Phi_{22}^\epsilon]$. The private-sector behavior is described in Φ form:

$$y'_{2,t} = x'_t \Phi_{\cdot 2} + \epsilon_{1,t} \Phi_{12}^\epsilon + \epsilon'_{2,t} \Phi_{22}^\epsilon. \quad (6)$$

In the remainder of this section we will extend the specification in (5) and (6) by allowing for censoring of $y_{1,t}$, censoring-regime-dependent coefficients for the private-sector equations, and stochastic volatility.

2.1 Censoring

In order to capture the ELB constraint, which we assume to be zero, we introduce censoring. We use $\tilde{y}_{1,t}^*$ to denote the desired or shadow interest rate, let $y_{1,t}^* = \tilde{y}_{1,t}^* - \mu_1$, and write the monetary policy rule as

$$y_{1,t}^* A_{11} + y'_{2,t} A_{21} = x'_t B_{\cdot 1} + \epsilon_{1,t}. \quad (7)$$

Here we replaced $y_{1,t}$ in (5) by $y_{1,t}^*$. The relationship between $y_{1,t}$ and $y_{1,t}^*$ is given by

$$y_{1,t} = \max \{y_{1,t}^*, -\mu_1\}.$$

The μ_1 threshold arises because $y_{1,t} = \tilde{y}_{1,t} - \mu_1$ and the ELB constraint applies to the actual and not the demeaned nominal interest rate. We will assume that both central bank and the private sector react to lagged $y_{1,t}$ instead of $y_{1,t}^*$. However, we do allow agents to react to the monetary policy shock when the ELB is binding. Both of these assumptions are consistent with the DSGE model in ACHSV. Thus, (6) remains unchanged. We define $y_t^{*'} = [y_{1,t}^*, y'_{2,t}]$.

2.2 Regime-Dependent Private-Sector Behavior

We will now extend the model and allow the private sector to change its behavior once the economy reaches the ELB. We introduce the observable regime (or ELB) indicator

$$s_t = \mathbb{I}\{y_{1,t} > -\mu_1\} \quad (8)$$

and write the private-sector equations as

$$y'_{2,t} = x'_t \Phi_{\cdot 2}(s_t) + u'_{2,t}(s_t) \quad (9)$$

where we define

$$u'_{2,t}(s_t) = \epsilon_{1,t} \Phi_{12}^\epsilon(s_t) + \epsilon'_{2,t} \Phi_{22}^\epsilon(s_t).$$

Plugging the expression for $y'_{2,t}$ from (9) into the monetary policy rule (7) leads to the Φ form of the interest rate equation:

$$y_{1,t}^* = \frac{1}{A_{11}} \left[x'_t (B_{\cdot 1} - \Phi_{\cdot 2}(s_t) A_{21}) + \epsilon_{1,t} (1 - \Phi_{12}^\epsilon(s_t) A_{21}) - \epsilon'_{2,t} \Phi_{22}^\epsilon(s_t) A_{21} \right]. \quad (10)$$

Define

$$\begin{aligned} \Phi_{\cdot 1}(s_t) &= \frac{1}{A_{11}} (B_{\cdot 1} - \Phi_{\cdot 2}(s_t) A_{21}), \\ u_{1,t}(s_t) &= \frac{1}{A_{11}} \left[\epsilon_{1,t} (1 - \Phi_{12}^\epsilon(s_t) A_{21}) - \epsilon'_{2,t} \Phi_{22}^\epsilon(s_t) A_{21} \right] \end{aligned} \quad (11)$$

such that we can write

$$y_{1,t}^* = x'_t \Phi_{\cdot 1}(s_t) + u_{1,t}(s_t). \quad (12)$$

In view of (8), for the model specification to be internally consistent, it has to be the case that whenever the “1” regression functions are active then $y_{1,t} = y_{1,t}^*$ and $y_{1,t}^* > -\mu_1$ must hold. Likewise, whenever the “0” regression functions are active, it has to be the case that $y_{1,t} = -\mu_1$ and $y_{1,t}^* \leq -\mu_1$. Given a set of parameters $A_{\cdot 1}$, $B_{\cdot 1}$, $\Phi_{\cdot 2}(s)$, $\Phi_{\cdot 2}^\epsilon(s)$, lagged values x_t , and a vector of structural shocks ϵ_t we can distinguish three cases:

Case 1 – Uniqueness: conditional on lagged values x_t and the innovation ϵ_t the state s_t , the latent variable $y_{1,t}^*$, and $y_{2,t}$ are uniquely determined. If

$$x'_t \Phi_{\cdot 1}(1) + u_{1,t}(1) > -\mu_1, \quad \text{then} \quad x'_t \Phi_{\cdot 1}(0) + u_{1,t}(0) > -\mu_1 \quad \text{which implies} \quad s_t = 1.$$

Alternatively, if

$$x'_t \Phi_{.1}(1) + u_{1,t}(1) \leq -\mu_1, \quad \text{then} \quad x'_t \Phi_{.1}(0) + u_{1,t}(0) \leq -\mu_1 \quad \text{which implies} \quad s_t = 0.$$

We use $\mathcal{E}^U(x_t)$ to denote the set of ϵ_t values for which s_t is unique.

Case 2 - Indeterminacy (Incompleteness): conditional on lagged values x_t and the innovation ϵ_t , the model is consistent with $s_t = 0$ and $s_t = 1$, which means that there are two possible values for $y_{1,t}^*$ and $y_{2,t}$, respectively. Formally,

$$x'_t \Phi_{.1}(1) + u_{1,t}(1) > -\mu_1 \quad \text{and} \quad x'_t \Phi_{.1}(0) + u_{1,t}(0) \leq -\mu_1 \quad \text{which implies} \quad s_t = 1 \text{ or } s_t = 0.$$

We use $\mathcal{E}^I(x_t)$ to denote the set of ϵ_t values for which s_t is not unique.

Case 3 - Non-existence (Incoherency): conditional on lagged values x_t and the innovation ϵ_t , the model is neither consistent with $s_t = 0$ nor $s_t = 1$ because

$$x'_t \Phi_{.1}(1) + u_{1,t}(1) \leq -\mu_1 \quad \text{and} \quad x'_t \Phi_{.1}(0) + u_{1,t}(0) > -\mu_1.$$

We use $\mathcal{E}^N(x_t)$ to denote the set of ϵ_t values for we have non-existence.

The fact that the existence and uniqueness of y_t depend on the lagged endogenous variables stacked in x_t and the structural innovations ϵ_t is an undesirable feature of the model. To rule out non-existence, in general one needs to restrict the domain of the innovations ϵ_t ; see Ascari and Mavroeidis (2020); Mavroeidis (2020). We will show in the following section that a necessary condition for the uniqueness condition to hold for all (x_t, ϵ_t) is that the piecewise linear private sector regression functions are continuous at the kink. Moreover, we provide a necessary and sufficient restriction on the VAR parameter space that guarantees uniqueness for all (x_t, ϵ_t) .

2.3 Piecewise Linear and Continuous Regression Functions

Building on ACHSV, we now impose that the private sector uses regression functions that are continuous at the kink. We refer to these regression functions as piecewise linear and continuous (PLC). Consider the monetary policy rule in Φ form. From (10) we deduce that the ELB starts to bind whenever

$$x'_t (B_{.1} - \Phi_{.2}(1)A_{21}) = -\mu_1 A_{11} - \epsilon_{1,t} (1 - \Phi_{12}^\epsilon(1)A_{21}) + \epsilon'_{2,t} \Phi_{22}^\epsilon(1)A_{21}. \quad (13)$$

Denote the j th element of the vector $x_{j,t}$, $j = 1, \dots, k$, with the understanding that $x_{k,t} = 1$ (intercept). Likewise, denote the j th rows of $B_{\cdot 1}$ and $\Phi_{\cdot 2}$ by B_{j1} and Φ_{j2} , respectively. Using this notation, we solve (13) for $x_{1,t}$ to obtain

$$\begin{aligned} x_{1,t} = & - \left(\sum_{j=2}^{k-1} x_{j,t} \frac{B_{j1} - \Phi_{j2}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right) - \frac{B_{k1} - \Phi_{k2}(1)A_{21} + \mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \\ & - \epsilon_{1,t} \frac{1 - \Phi_{12}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} + \epsilon'_{2,t} \frac{\Phi_{22}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}}. \end{aligned} \quad (14)$$

Continuity at the kink implies that

$$\begin{aligned} x_{1,t} \Phi_{12}(1) + \left(\sum_{j=2}^{k-1} x_{j,t} \Phi_{j,2}(1) \right) + \Phi_{k2}(1) + \epsilon_{1,t} \Phi_{12}^\epsilon(1) + \epsilon'_{2,t} \Phi_{22}^\epsilon(1) \\ = x_{1,t} \Phi_{12}(0) + \left(\sum_{j=2}^{k-1} x'_{j,t} \Phi_{j,2}(0) \right) + \Phi_{k2}(0) + \epsilon_{1,t} \Phi_{12}^\epsilon(0) + \epsilon'_{2,t} \Phi_{22}^\epsilon(0). \end{aligned} \quad (15)$$

Now plug the expression for $x_{1,t}$ in (14) into Equation (15) and use the continuity restrictions to solve for the coefficients in the $s = 0$ regime:

$$\begin{aligned} \Phi_{j2}(0) &= \Phi_{j2}(1) + \frac{B_{j1} - \Phi_{j2}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta, \quad j = 2, \dots, k-1 \\ \Phi_{k2}(0) &= \Phi_{k2}(1) + \frac{B_{k1} - \Phi_{k2}(1)A_{21} + \mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta \\ \Phi_{12}^\epsilon(0) &= \Phi_{12}^\epsilon(1) + \frac{1 - \Phi_{12}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta \\ \Phi_{22}^\epsilon(0) &= \Phi_{22}^\epsilon(1) - \frac{\Phi_{22}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta, \end{aligned} \quad (16)$$

where

$$\Phi_{12}^\Delta = \Phi_{12}(0) - \Phi_{12}(1).$$

The unrestricted coefficient matrices are $\Phi_{\cdot 2}(1)$, $\Phi_{\cdot 2}^\epsilon(1)$, and Φ_{12}^Δ . The dimension of Φ_{12}^Δ is $(n-1) \times 1$. Two special cases are noteworthy. First, if $\Phi_{12}^\Delta = 0$ then the regression functions are strictly linear and have no kink: $\Phi_{\cdot 2}(0) = \Phi_{\cdot 2}(1)$ and $\Phi_{\cdot 2}^\epsilon(0) = \Phi_{\cdot 2}^\epsilon(1)$. Second, if the private sector does not react to the monetary policy shock, i.e., $\Phi_{12}^\epsilon(1) = 0$ and $\Phi_{12}^\epsilon(0) = 0$, then $\Phi_{12}^\Delta = 0$, which in turn implies that there is no kink in the reactions to $x_{2,t}$ and $\epsilon_{2,t}$: $\Phi_{22}(0) = \Phi_{22}(1)$ and $\Phi_{22}^\epsilon(0) = \Phi_{22}^\epsilon(1)$.

2.4 PLC Regression Functions and Uniqueness

The following proposition summarizes our theoretical results. First, we show that continuity of the regression functions is a necessary condition to achieve uniqueness without having to restrict the domain of ϵ_t conditional on the lagged values x_t . Second, we provide necessary and sufficient restrictions for the VAR parameters that ensure uniqueness for all (x_t, ϵ_t) .

Proposition 1

(i) *(Necessary Condition)* Only if the piecewise linear private sector regression functions are continuous at the kink, then uniqueness can be obtained for all (x_t, ϵ_t) , i.e., $\mathbb{P}_{x_t}\{\epsilon_t \in \mathcal{E}^U(x_t)\} = 1$ for all x_t .

(ii) *(Necessary and Sufficient Condition)* Suppose that the piecewise linear private sector regression functions are continuous at the kink. If and only if

$$\text{sign}(B_{11} - \Phi_{12}(1)A_{21}) = \text{sign}(B_{11} - \Phi_{12}(0)A_{21}),$$

then uniqueness is obtained for all (x_t, ϵ_t) , i.e., $\mathbb{P}_{x_t}\{\epsilon_t \in \mathcal{E}^U(x_t)\} = 1$ for all x_t .

A proof of the proposition is provided in the Online Appendix. In our estimation, we will impose the parameter restrictions described in Proposition 1(ii) to ensure uniqueness of $(s_t, y_{1,t}^*, y_{2,t})$ conditional on all values of (x_t, ϵ_t) .

2.5 Comparison to Mavroeidis (2020)

Our model is related to the censored VAR specification considered in Mavroeidis (2020). He writes the private-sector equations in A form, which can be stated as

$$y_{1,t}^*A_{12}^* + y_{1,t}\tilde{A}_{12} + y'_{2,t}A_{22} = x'_tB_{.2} + \epsilon'_{2,t}, \quad y_{1,t} = \max\{y_{1,t}^*, -\mu_1\}.$$

This specification allows agents to react to a linear combination of the shadow rate $y_{1,t}^*$ and the actual (constrained) interest rate $y_{1,t}$. Using our state-dependent notation, his private sector specification can be rewritten as

$$y_{1,t}^*A_{12}(s_t) + y'_{2,t}A_{22} = x'_tB_{.2}(s_t) + \epsilon'_{2,t}, \tag{17}$$

where

$$s_t = \mathbb{I}\{y_{1,t} > -\mu_1\}, \quad A_{12}(s) = \begin{cases} A_{12}^* + \tilde{A}_{12} & \text{if } s = 1 \\ A_{12}^* & \text{if } s = 0 \end{cases}$$

$$B_{\cdot 2}(s) = [B_{12} | \dots | B_{(k-1)2} | B_{k2}(s)], \quad B_{k2}(s) = \begin{cases} 0 & \text{if } s = 1 \\ \mu_1(A_{12}(1) - A_{12}(0)) & \text{if } s = 0 \end{cases}.$$

Here B_{j2} , $j = 1, \dots, k$, is the j th row of $B_{\cdot 2}$. Note that only the last row of $B_{\cdot 2}$ is regime-dependent. Recall that we fixed it to zero for $s = 1$. For $s = 0$ it captures that $y_{1,t} = -\mu_1$ in the ELB regime.⁵

We will refer to our model, comprising (7) and (9), as model $\mathcal{M}(\Phi_{\cdot 2}(s), \Phi_{\cdot 2}^\epsilon(s))$, and to the (restricted) Mavroeidis model, comprising (7) and (17) as model $\mathcal{M}(A_{12}(s))$ model. Moreover, we denote the $\mathcal{M}(\Phi_{\cdot 2}(s), \Phi_{\cdot 2}^\epsilon(s))$ with PLC private sector regression functions as $\mathcal{M}_{PLC}(\Phi_{\cdot 2}(s), \Phi_{\cdot 2}^\epsilon(s))$. The relationship between three specifications is summarized in the following proposition.⁶

Proposition 2

- (i) $\mathcal{M}(\Phi_{\cdot 2}(s), \Phi_{\cdot 2}^\epsilon(s)) \supset \mathcal{M}(A_{12}(s))$,
- (ii) $\mathcal{M}_{PLC}(\Phi_{\cdot 2}(s), \Phi_{\cdot 2}^\epsilon(s)) = \mathcal{M}(A_{12}(s))$.

A proof of the proposition is provided in the Online Appendix. The intuition for Part (i) is straightforward. We started out in this paper from a very general model specification in which the private-sector equations are regime dependent. The possibility of private-sector behavior changing when the ELB becomes binding generates an additional $(n-1) \times (k+n)$ coefficients, compared to a constant-coefficient VAR. The $\mathcal{M}(A_{12}(s))$ model, on the other hand, only adds $(n-1)$ coefficients.

Part (ii) of the proposition combines two insights. First, imposing a PLC structure onto the private-sector equations reduces the number of additional unrestricted parameters in the ELB regime from $(n-1) \times (k+n)$ to $n-1$ parameters, see (16), which is the same number of additional parameters as in the $\mathcal{M}(A_{12}(s))$ model. Second, with some algebra it can be shown that $\mathcal{M}(A_{12}(s))$ delivers PLC regression functions for the private sector. The equality

⁵Mavroeidis (2020) refers to the model with $A_{12}^* = 0$ as kinked SVAR (KSVAR), to the model with $\tilde{A}_{12} = 0$ as censored SVAR (CSVAR), and the general specification as censored and kinked SVAR (CKSVAR).

⁶We are grateful to one of the referees for providing valuable conjectures that lead to Proposition 2.

of the model sets implies that PLC regression functions for the private sector can only be obtained if the state-dependence in the A representation of the private sector equations is concentrated in the reaction to shadow rate movements.

We show in the Online Appendix that by solving the locus at which the ELB becomes binding (13) for $\epsilon_{1,t}$ instead of $x_{1,t}$, the necessary and sufficient condition in Proposition 1(ii) can equivalently be expressed as

$$\text{sign}(1 - \Phi_{12}^\epsilon(1)A_{21}) = \text{sign}(1 - \Phi_{12}^\epsilon(0)A_{21}).$$

Moreover, we show that the condition is equivalent to the condition stated in Proposition 1 of Mavroeidis (2020), which in turn dates back to Gourieroux, Laffont, and Monfort (1980). This is unsurprising given our Proposition 2(ii).

As mentioned in the introduction, our SVAR specification allows for heteroskedastic structural shocks, which is important for inference on switches in the conditional mean function. Mavroeidis' specification allows for both lags of $y_{1,t}$ and the latent variable $y_{1,t}^*$ on the right-hand side of the VAR specification, whereas we stay close to the DSGE model specification in ACHSV and only use the former.

2.6 Stochastic Volatility

In order to make the empirical model more flexible, we allow for stochastic volatility in the structural innovations ϵ_t . We replace the homoskedasticity assumption $\epsilon_t \sim N(0, D)$ by

$$\epsilon_t \sim N(0, D_t), \quad \ln D_{ii,t} = \ln D_{ii} + \zeta_{i,t}, \quad \zeta_{i,t} = \rho_{\zeta,i}\zeta_{i,t-1} + \varsigma_i\eta_{i,t}, \quad \eta_{i,t} \sim N(0, 1). \quad (18)$$

Here $D_{ii,t}$, $i = 1, \dots, n$ are the n diagonal elements of the (diagonal) matrix D_t . For each variable i we obtain two additional parameters, $(\rho_{\zeta,i}, \varsigma_i)$ that govern the stochastic volatility process. The parameter $\rho_{\zeta,i}$ controls the persistence and ς_i the variance of the stochastic volatility process. (2) is replaced by an equation with time-dependent covariance matrices:

$$u_t = \Phi_\epsilon(s_t)\epsilon_t, \quad u_t \sim N(0, \Sigma_t(s_t)) \text{ where } \Sigma_t(s_t) = \Phi_\epsilon(s_t)D_t\Phi_\epsilon'(s_t). \quad (19)$$

Note that none of the calculations in Section 2.3 depended on the covariance matrices D and Σ . Thus, the continuity restrictions in (16) and the result in Proposition 1 are not affected by the presence of heteroskedasticity.

2.7 Parameter Summary

After allowing for regime-dependent private-sector regression functions and imposing continuity at the kink, the parameters of the SVAR model with stochastic volatility are:

$$\underbrace{\mu}_{n \times 1}, \underbrace{A_{\cdot 1}}_{n \times 1}, \underbrace{B_{\cdot 1}}_{k \times 1}, \underbrace{\Phi_{\cdot 2}(1)}_{k \times (n-1)}, \underbrace{\Phi_{\cdot 2}^{\epsilon}(1)}_{n \times (n-1)}, \underbrace{\Phi_{12}^{\Delta}}_{1 \times (n-1)}, \underbrace{\{D_{ii}\}_{i=1}^n}_n, \underbrace{\{\rho_{\zeta,i}, \varsigma_i\}_{i=1}^n}_{2n}. \quad (20)$$

The interest rate coefficient in $A_{\cdot 1}$ is normalized to one. Moreover, the elements in the last rows of $B_{\cdot 1}$ and $\Phi_{\cdot 2}(1)$ are fixed at zero because the level of the observables \tilde{y}_t is captured by μ . Notationally, it is convenient to keep the intercepts in the specification, because they are non-zero for $\Phi_{\cdot 2}(0)$ in the ELB regime.

3 Prior Distribution

In our subsequent empirical application we assume that $y_{1,t}$ corresponds to the nominal interest rate R_t , and $y'_{2,t} = [z_t, \pi_t]$, where z_t is output gap and π_t is inflation. We interpret the innovations $\epsilon_t = [\epsilon_{R,t}, \epsilon_{D,t}, \epsilon_{S,t}]$ as monetary policy, demand, and supply shocks, respectively. We use Bayesian techniques to estimate the SVAR with ELB censoring and piecewise-linear regression functions. This requires the specification of a prior distribution. Rather than specifying the prior directly on the unrestricted elements of the parameters listed in (20), we consider several transformations to facilitate the elicitation of the prior. These transformations are presented in Section 3.1. Section 3.2 summarizes the specification of our baseline prior. Section 3.3 discusses adjustments to the baseline prior and the treatment of hyperparameters.

3.1 (Re)parameterization of the SVAR

We follow Baumeister and Hamilton (2018) and specify a prior on contemporaneous interactions between interest rates, output gap, and inflation in the regime in which the ELB is not binding. Using the notation in (4), these interactions are summarized in the matrix $A(1)$, which we partition into $A(1) = [A_{\cdot 1}, A_{\cdot 2}(1)]$. The first column of $A(1)$, denoted by $A_{\cdot 1}$, is included in the parameter list (20) and does not depend on the state s . The remaining $n-1$ columns, collected in the matrix $A_{\cdot 2}(1)$, replace the $n \times (n-1)$ matrix $\Phi_{\cdot 2}^{\epsilon}(1)$ in (20).

To obtain an expression for $A_{.2}(s)$ as a function of $\Phi_{.2}^\epsilon(s)$, we solve (5) for $\epsilon_{1,t}$ and plug the expression into (9). This yields:

$$y'_{2,t} = x'_t \Phi_{.2}(s) + (y_{1,t} A_{11} + y'_{2,t} A_{21} - x'_t B_{.1}) \Phi_{12}^\epsilon(s) + \epsilon'_{2,t} \Phi_{22}^\epsilon(s).$$

After re-arranging terms we obtain

$$-y_{1,t} A_{11} \Phi_{12}^\epsilon(s) + y'_{2,t} (I - A_{21} \Phi_{12}^\epsilon(s)) = x'_t (\Phi_{.2}(s) - B_{.1} \Phi_{12}^\epsilon(s)) + \epsilon'_{2,t} \Phi_{22}^\epsilon(s). \quad (21)$$

Let $A_{.2}(s) = [A_{12}(s) | A_{22}(s)]$ and deduce that

$$\underbrace{A_{12}(s)}_{1 \times (n-1)} = -A_{11} \Phi_{12}^\epsilon(s) (\Phi_{22}^\epsilon(s))^{-1}, \quad \underbrace{A_{22}(s)}_{(n-1) \times (n-1)} = (I - A_{21} \Phi_{12}^\epsilon(s)) (\Phi_{22}^\epsilon(s))^{-1}. \quad (22)$$

We follow Baumeister and Hamilton (2018) by expressing the matrix of contemporaneous interactions, in our case for the non-binding regime $s = 1$, as a function of the parameters $[\psi_\pi, \psi_z, \alpha_S, \beta_D, \gamma_D]'$:

$$A(1) = \begin{bmatrix} 1 & -\gamma_D & 0 \\ -\psi_z & 1 & 1 \\ -\psi_\pi & -\beta_D & -\alpha_S \end{bmatrix}. \quad (23)$$

The parameterization of the $A(1)$ matrix is motivated by a three-equation New Keynesian DSGE model. Each column corresponds to an equation, and each row to a variable. The “1” entries are normalizations. Recall that the observables are ordered as follows: $y_t = [R_t, z_t, \pi_t]$. We interpret the first column of $A(1)$ as a Taylor rule with output and inflation gap coefficients ψ_z and ψ_π , respectively.

The second column represents the Euler equation after the expectations of next period’s inflation and output gap have been replaced by AR(1) forecasts. This creates a contemporaneous relationship between the output gap (being proportional to consumption), the nominal interest rate, and inflation. The parameter β_D can be interpreted as an aggregate demand elasticity. The third column represents the New Keynesian Phillips curve with marginal costs replaced by output gap and expected inflation replaced by an AR(1) forecast. The coefficient α_S can be interpreted as a supply elasticity.

We replace α_S by $\alpha_S^\Delta = \alpha_S - \beta_D$ and impose the following domain/sign restrictions on the coefficients:

$$\psi_\pi \geq 0, \quad \psi_z \geq 0, \quad \alpha_S^\Delta \geq 0, \quad \beta_D \geq 0, \quad \gamma_D \leq 0. \quad (24)$$

The restriction of $\beta_D \geq 0$ may seem counterintuitive in view of the interpretation of the coefficient as a demand elasticity. However, it is consistent with a DSGE model in which the demand equation represents a consumption Euler equation and a rise in (expected) inflation lowers real returns which creates an incentive to increase current-period consumption.

To facilitate the elicitation of a prior distribution, it is helpful to derive the inverse of $A(1)$, which determines the contemporaneous impact of the three innovations:

$$[A(1)]^{-1} \propto \begin{bmatrix} \alpha_S^\Delta & (\beta_D + \alpha_S^\Delta)\gamma_D & \gamma_D \\ \psi_\pi + (\beta_D + \alpha_S^\Delta)\psi_z & (\beta_D + \alpha_S^\Delta) & 1 \\ -(\psi_\pi + \beta_D\psi_z) & -(\beta_D + \gamma_D\psi_\pi) & \gamma_D\psi_z - 1 \end{bmatrix}. \quad (25)$$

Here \propto denotes proportionality. The first row of the matrix captures the response of interest rates, the output gap, and inflation to a monetary policy shock $\epsilon_{1,t}$. In order for interest rates to rise, and output and inflation to fall in response to a contractionary monetary policy shock we need $\alpha_S^\Delta \geq 0$, $(\beta_D + \alpha_S^\Delta) \geq 0$, and $\gamma_D \leq 0$, all of which are implied by (24). The sum $(\beta_D + \alpha_S^\Delta)$ controls the relative response of output gap and inflation and the ratio γ_D/α_S^Δ is the relative response of interest rates and inflation.

The second row of $[A(1)]^{-1}$ represents the contemporaneous demand shock response. Output gap and prices move in the same direction because we imposed $\beta_D + \alpha_S^\Delta \geq 0$ and the policy rule coefficients ψ_π and ψ_z are also positive. Finally, the third row of the matrix in (25) determines the response to a supply shock. The directions of the inflation and output gap responses to a supply shock depend on the specific parameterization. For instance, assuming that $\gamma_D < 0$, inflation is falling in response to the supply shock. The output response is given by $-(\beta_D + \gamma_D\psi_\pi)$. We observe a rise in output if $-\gamma_D > \beta_D/\psi_\pi$.

Our prior distribution is specified over the set of parameters

$$\theta = (\mu, \psi_\pi, \psi_z, \alpha_S^\Delta, \beta_D, \gamma_D, B_{\cdot 1}, \{\Phi_{\cdot i}(1)\}_{i=2}^n, \Phi_{12}^\Delta, \{\sqrt{D_{ii}}\}_{i=1}^n, \{\rho_{\zeta, i}, \varsigma_i\}_{i=1}^n). \quad (26)$$

3.2 Distributional Assumptions

Define the vector of hyperparameters $\lambda = [\lambda_I, \lambda_\phi, \lambda_d, \lambda_0, \lambda_l, \lambda_\Delta]'$. We consider a baseline prior of the form

$$\begin{aligned} \tilde{p}(\theta|\lambda) &= \tilde{p}(\mu|\lambda_I)\tilde{p}(\vartheta) \left(\prod_{i=1}^n \tilde{p}(\sqrt{D_{ii}}|\vartheta; \lambda_d)\tilde{p}(B_{\cdot 1}|D; \lambda_\phi, \lambda_0, \lambda_l)\tilde{p}(\rho_{\zeta, i}, \varsigma_i) \right) \\ &\times \left(\prod_{i=2}^n \tilde{p}(\Phi_{\cdot i}(1)|\vartheta, D; \lambda_\phi, \lambda_0, \lambda_l) \right) \tilde{p}(\Phi_{12}^\Delta|\lambda_\Delta), \text{ where } \vartheta = (\psi_\pi, \psi_z, \alpha_S^\Delta, \beta_D, \gamma_D). \end{aligned} \quad (27)$$

Recall that D is the diagonal matrix with elements D_{ii} , $i = 1, \dots, n$. As previously noted in Section 2.7, the last element of the vectors $B_{\cdot 1}$ and $\Phi_{\cdot i}(1)$, respectively, is set equal to zero and essentially excluded from the construction of the prior density.

Prior for Level Parameters $\tilde{p}(\mu|\lambda_I)$. We use independent normal distributions of the form $\mu_i \sim N(\underline{\mu}_i, \underline{s}_i^2/\lambda_I)$, $i = 1, \dots, n$. In the empirical application we set $\underline{\mu}_i$ to the pre-ELB sample mean of $y_{i,t}$ computed over the period from 1984:Q1 to 2006:Q4, and \underline{s}_i is its sample standard deviation. The hyperparameter λ_I scales the prior precision of the μ_i s.

Prior for Contemporaneous Interactions $\tilde{p}(\psi_\pi, \psi_z, \alpha_S^\Delta, \beta_D, \gamma_D)$. We assume that the underlying parameters are independent:

$$\tilde{p}(\psi_\pi, \psi_z, \alpha_S^\Delta, \beta_D, \gamma_D) = \tilde{p}(\psi_\pi)\tilde{p}(\psi_z)\tilde{p}(\alpha_S^\Delta)\tilde{p}(\beta_D)\tilde{p}(\gamma_D).$$

The first section of Table 1 summarizes the specification of the baseline prior for the $A(1)$ coefficients. The distributions are chosen so that the sign restrictions in (24) hold. The priors for (ψ_π, ψ_z) are broadly in line with priors used in the DSGE model literature.⁷ The numerical values for $(\alpha_S^\Delta, \beta_D, -\gamma_D)$ are more difficult to interpret. In the empirical section we will plot the prior distribution of impulse responses.

Prior for Innovation Variances $\tilde{p}(\sqrt{D_{ii}}|\vartheta; \lambda_d)$. We specify a prior for $\sqrt{D_{ii}}$, $i = 1, \dots, n$. Starting point is a prior for the D_{ii} elements, which takes the form of an Inverse Gamma distribution that is parameterized as scaled Inverse χ^2 distribution with density

$$\tilde{p}(D_{ii}|\underline{S}_{ii}, \lambda_d) = \frac{[(\lambda_d/2)\underline{S}_{ii}]^{\lambda_d/2}}{\Gamma(\lambda_d/2)} D_{ii}^{-\lambda_d/2-1} \exp\{-(\lambda_d/2)\underline{S}_{ii}D_{ii}^{-1}\}.$$

⁷In a DSGE model, the monetary policy rule is typically parameterized in terms of $\psi_\pi = (1 - \rho)\tilde{\psi}_\pi$ and $\psi_z = (1 - \rho)\tilde{\psi}_z$, where $\rho \approx 0.8$ is an interest rate smoothing coefficient.

Table 1: Baseline Prior Distribution for Various Coefficients

Param.	Distr.	P(1)	P(2)
ψ_π	Gamma	1.5	0.5
ψ_z	Gamma	0.5	0.25
α_S^Δ	Gamma	1.0	0.7
β_D	Gamma	1.0	0.7
$-\gamma_D$	Gamma	0.25	0.12
$\rho_{\zeta,i}$	Beta	0.75	0.10
ς_i	InvGamma	0.1	4.0
λ_ϕ	Beta	0.75	0.10
λ_0	Gamma	10.0	5.00
λ_l	Gamma	2.00	1.00

Notes: The table lists marginal distributions. P(1) and P(2) are means and standard deviations for Beta and Gamma distributions. The Inverse Gamma distribution is parameterized as scaled inverse χ^2 distribution with density $p(\zeta^2|s^2, \nu) \propto (\sigma^2)^{-\nu/2-1} \exp[-\nu s^2/(2\sigma^2)]$, where P(1) is $\sqrt{s^2}$ and P(2) is ν . The density of ζ is obtained by the change of variables $\sigma = \sqrt{\sigma^2}$. The following hyperparameters are fixed: $\lambda_l = \infty$, $\lambda_d = 4$, and $\lambda_\Delta = 16$.

The density of $\sqrt{D_{ii}}$ is obtained by a change of variable, which adds the Jacobian term $2\sqrt{D_{ii}}$ to the density. Here, λ_d is a hyperparameter that controls the degrees of freedom of the scaled Inverse χ^2 distribution. We use an estimate of the variance of $y_t' A(1)$ to choose \underline{S}_{ii} . Specifically, we fit univariate AR(p) models to the series $y_{i,t}$ and let s_i^2 be the estimated innovation variance. We use ϑ to generate the $A(1)$ matrix in (23) and define the matrix

$$\underline{S} = A(1)' \begin{bmatrix} s_1^2 & 0 & 0 \\ 0 & s_2^2 & 0 \\ 0 & 0 & s_3^2 \end{bmatrix} A(1).$$

The values \underline{S}_{ii} correspond to the diagonal elements of \underline{S} .

Prior for the Stochastic Volatility Processes $\tilde{p}(\rho_{\zeta,i})\tilde{p}(\varsigma_i)$. We assume that the parameters are independently and identically distributed across i . We use a Beta distribution for $\rho_{\zeta,i}$ and an Inverse Gamma distribution for ς_i . The parameterizations of these distributions are listed in Table 1.

Prior for Policy Rule Coefficients $\tilde{p}(B_{\cdot 1}|D; \lambda_\phi, \lambda_0, \lambda_l)$. We will now specify a prior for $B_{\cdot 1}$. Let

$$B_{\cdot 1}|(D, \lambda_\phi, \lambda_0, \lambda_l) \sim N(\underline{B}_{\cdot 1}(\lambda_\phi), \underline{P}_{\cdot 1}^{-1}(\lambda_0, \lambda_l)). \quad (28)$$

We set the prior mean of B_{11} , i.e., the coefficient on R_{t-1} , equal to $\underline{B}_{11} = \lambda_\phi$. The prior

mean for the remaining coefficients is set to zero. The $k \times k$ precision matrix $\underline{P}_{\cdot 1}(\lambda_0, \lambda_l, \lambda_I)$ is diagonal. The value associated with the coefficient of the l th lag of variable j is $\lambda_0 l^{\lambda_l} s_j^2 / D_{11}$, where s_j was defined above and D_{11} is the innovation variance in the interest rate equation. Here λ_0 is a hyperparameter that scales the overall precision of the prior, and λ_l determines how quickly the prior variance decays with lag length l . The prior precision of the last element of $B_{\cdot 1}$ (intercept) is infinite, because we fixed it at zero.⁸

Prior for Private-Sector Coefficients $\tilde{p}(\Phi_{\cdot i}(1) | \vartheta, D; \lambda_\phi, \lambda_0, \lambda_l)$. We will now specify a prior for the $k \times 1$ column vectors $\Phi_{\cdot i}(1)$ that stack the reduced-form coefficients for private-sector variable i , where $i = 2, \dots, n$. From (4) we deduce that the reduced-form forecast errors are given by

$$u'_t = \epsilon'_t [A(1)]^{-1},$$

as before, the $A(1)$ matrix is generated from ϑ . The covariance matrix of the forecast errors is

$$\Sigma(A, D) = [A(1)]^{-1'} \mathbb{E}[\epsilon_t \epsilon'_t] [A(1)]^{-1} = [A(1)]^{-1'} D [A(1)]^{-1}.$$

Let

$$\Phi_{\cdot i}(1) | (D; \lambda_\phi, \lambda_0, \lambda_l) \sim N(\underline{\Phi}_{\cdot i}(\lambda_\phi), \underline{P}_i^{-1}(\lambda_0, \lambda_l)), \quad i = 2, \dots, n. \quad (29)$$

In order to impose the belief that the individual series are well approximated by AR(1) processes we set the prior mean vector as follows: $\underline{\Phi}_{ji}(\lambda_\phi) = \lambda_\phi$ for $j = i$ – this coefficient interacts with $y_{i,t-1}$ – and $\underline{\Phi}_{ji}(\lambda_\phi) = 0$ for $j \neq i$, where $j = 1, \dots, k$. The $k \times k$ precision matrix $\underline{P}_i(\lambda_0, \lambda_l, \lambda_I)$ is diagonal. The value associated with the coefficient of the l th lag of variable j is $\lambda_0 l^{\lambda_l} s_j^2 / \Sigma_{ii}$, where s_j was defined above. As before, the hyperparameter λ_0 scales the overall precision of the prior, and λ_l determines how quickly the prior variance decays with lag length l . The last element, $\Phi_{ki}(1)$, capturing the intercept is fixed at zero.

Prior for Regression Function Differentials $p(\Phi_{12}^\Delta)$. We assume that

$$\Phi_{12}^\Delta | \lambda_\Delta \sim N\left(0, \frac{1}{\lambda_\Delta} I\right). \quad (30)$$

Note that for $\Phi_{12}^\Delta = 0$ the regression functions remain unchanged once the economy reaches the ELB.

⁸In practice, we are reducing the dimension of $\underline{P}_{\cdot 1}(\lambda_0, \lambda_l, \lambda_I)$ to $(k-1) \times (k-1)$.

3.3 Adjustments to the Baseline Prior and Hyperparameters

We make two adjustments to the prior $\tilde{p}(\theta|\lambda)$ in (27). These adjustments are implemented as follows:

$$p(\theta|\lambda) \propto \tilde{p}(\theta|\lambda)f_u(\theta)f_s(\theta). \quad (31)$$

Uniqueness. We enforce the uniqueness restriction in Proposition 1 and define the indicator function

$$f_u(\theta) = \mathbb{I}\left\{\text{sign}(B_{11} - \Phi_{12}(1)A_{21}) = \text{sign}(B_{11} - \Phi_{12}(0)A_{21})\right\}.$$

Stationarity. We impose that the reduced form representation of the SVAR is stationary conditional on $s_t = 1$ and $s_t = 0$ for all t , respectively. To do so, we convert the $A\Phi$ representation of the SVAR into the Φ representation in (1) and check that all roots of the characteristic polynomial $I - \sum_{l=1}^p \Phi_l(s)z^l$ are outside of the unit circle for $s = 1$ and $s = 0$. Let $f_s(\theta)$ be the indicator function that is equal to one if the stationarity condition is satisfied and equal to zero otherwise.

Hyperparameters. We adopt a hierarchical modeling approach and specify a prior for the hyperparameters as well. Giannone, Lenza, and Primiceri (2015) showed that this approach leads to a good empirical fit and forecasting performance for Bayesian VARs. Let $\tilde{p}(\lambda)$ be a properly normalized density. Then we define

$$p(\theta, \lambda) \propto \tilde{p}(\theta|\lambda)f_u(\theta)f_s(\theta)\tilde{p}(\lambda). \quad (32)$$

Under this construction, the conditional prior distribution remains equal to $p(\theta|\lambda)$, as specified in (31). The marginal prior of λ is given by

$$p(\lambda) \propto \tilde{p}(\lambda) \int \tilde{p}(\theta|\lambda)f_u(\theta)f_s(\theta)d\theta. \quad (33)$$

The integral on the right-hand side re-weights the density $\tilde{p}(\lambda)$. This prior specification avoids having to evaluate the integral as part of the posterior sampling and has the advantage that it down-weights values of λ under which the uniqueness and stationarity conditions are violated with very high probability. Prior distributions for λ_ϕ , λ_0 , and λ_l are reported in Table 1. The remaining hyperparameters are fixed at $\lambda_I = \infty$ (we are fixing the level parameters μ), $\lambda_d = 4$, and $\lambda_\Delta = 16$.

4 Likelihood Function

We now derive the likelihood associated with the SVAR model discussed in Section 2 under the assumption that the uniqueness condition in Proposition 1 is satisfied. In Section 4.1 we first discuss how to evaluate the exact likelihood function if the innovations are homoskedastic. In Section 4.2 we show how a particle filter approximation of the likelihood function can be obtained for the model with heteroskedasticity.

4.1 Homoskedasticity

We factorize the likelihood function as follows

$$p(Y_{1:T}|\theta) = \prod_{t=1}^T p(y_{1,t}|Y_{1:t-1}, \theta)p(y_{2,t}|y_{1,t}, Y_{1:t-1}, \theta), \quad (34)$$

where $Y_{t_1:t_2}$ denotes the sequence y_{t_1}, \dots, y_{t_2} .

Parameter Transformations. We begin with several parameter transformations to obtain the Φ representation of the SVAR. Based on θ we can compute

$$A_{\cdot 1}, A_{\cdot 2}(1), B_{\cdot 1}, \Phi_{\cdot 2}(1), \Phi_{12}^{\Delta}, D.$$

From $A(1)$ we obtain $\Phi^{\epsilon}(1) = [A(1)]^{-1}$. This leads to the Φ form for $y_{2,t}$ when $s_t = 1$:

$$y_{2,t} = x_t' \Phi_{\cdot 2}(1) + \epsilon_{1,t} \Phi_{12}^{\epsilon}(1) + \epsilon_{2,t}' \Phi_{22}^{\epsilon}(1). \quad (35)$$

We proceed by transforming the monetary policy rule so that we obtain the Φ form for $y_{1,t}$. Plugging the expression for $y_{2,t}$ in (35) into the monetary policy rule (5) we obtain:

$$y_{1,t} = \frac{1}{A_{11}} \left[x_t' (B_{\cdot 1} - \Phi_{\cdot 2}(1)A_{21}) + \epsilon_{1,t} (1 - \Phi_{12}^{\epsilon}(1)A_{21}) - \epsilon_{2,t}' \Phi_{22}^{\epsilon}(1)A_{21} \right] \quad (36)$$

We deduce that

$$\Phi_{\cdot 1}(1) = \frac{1}{A_{11}} (B_{\cdot 1} - \Phi_{\cdot 2}(1)A_{21}), \quad \Sigma(1) = \Phi^{\epsilon}(1)D\Phi^{\epsilon'}(1). \quad (37)$$

To obtain the Φ form for $s_t = 0$, let $\Phi_{12}(0) = \Phi_{12}(1) + \Phi_{12}^{\Delta}$ and use (16) to compute $\Phi_{\cdot 2}(0)$ and $\Phi_{22}^{\epsilon}(0)$. Then follow the steps in (35) to (37) to obtain $\Phi(0)$ and $\Sigma(0)$.

Period- t Densities. We partition the covariance matrix $\Sigma(s)$ into $[\Sigma_{ij}(s)]$ such that the partitions conform with $y_t = [y_{1,t}, y'_{2,t}]'$ and define

$$\Sigma_{2|1}(s) = \Sigma_{22}(s) - \Sigma_{21}(s)\Sigma_{11}^{-1}(s)\Sigma_{12}(s), \quad M_{12}(s) = \Sigma_{11}^{-1}(s)\Sigma_{12}(s), \quad M_{21}(s) = M'_{12}(s).$$

Moreover, let

$$u_{1,t}(s_t) = y_{1,t} - x'_t\Phi_{\cdot 1}(s_t) \quad u'_{2,t}(s_t) = y'_{2,t} - x'_t\Phi_{\cdot 2}(s_t).$$

The density $p(y_{1,t}|\cdot)$ is obtained as follows. If the conditions in Proposition 1 are satisfied, then $y_{1,t} = -\mu_1$ whenever $u_{1,t}(1) \leq -x'_t\Phi_{\cdot 1}(1) - \mu_1$. Thus, the distribution of $y_{1,t}$ is a mixture of a pointmass at $-\mu_1$ and a continuous distribution with support on $(-\mu_1, \infty)$. We write its density as

$$p(y_{1,t}|\cdot) = \mathbb{I}\{y_{1,t} = -\mu_1\}F_N\left(-\frac{x'_t\Phi_{\cdot 1}(1) + \mu_1}{\sqrt{\Sigma_{11}(1)}}\right) + \mathbb{I}\{y_{1,t} > -\mu_1\}p_N(u_{1,t}(1); 0, \Sigma_{11}(1)), \quad (38)$$

where $F_N(\cdot)$ is the cumulative distribution function of a $N(0, 1)$ and $p_N(x; \mu, \Sigma)$ is the probability density function of a $N(\mu, \Sigma)$. The derivation of $p(y_{2,t}|y_{1,t}, \cdot)$ is more tedious and relegated to the Online Appendix. We have to distinguish between $y_{1,t} = -\mu_1$ and $y_{1,t} > -\mu_1$. The densities are given by

$$p(y_{2,t}|y_{1,t} > -\mu_1, \cdot) = p_N(u_{2,t}(1); u_{1,t}M_{12}(1), \Sigma_{2|1}(1)). \quad (39)$$

and

$$\begin{aligned} & p(y_{2,t}|y_{1,t} = -\mu_1, \cdot) \quad (40) \\ &= (2\pi)^{-(n-1)/2}|\Sigma_{2|1}(0)|^{-1/2}|\Sigma_{11}(0)|^{-1/2}|\bar{V}_u(0)|^{1/2} \left[F_N\left(-\frac{x'_t\Phi_{\cdot 1}(1) + \mu_1}{\sqrt{\Sigma_{11}(1)}}\right) \right]^{-1} \\ & \times F_N\left(-\frac{x'_t\Phi_{\cdot 1}(0) + \mu_1 + \bar{M}_u(0)u_{2,t}(0)}{\sqrt{\bar{V}_u(0)}}\right) \\ & \times \exp\left\{-\frac{1}{2}u'_{2,t}(0)\Sigma_{2|1}^{-1}(0)u_{2,t}(0) + \frac{1}{2}\bar{V}_u^{-1}(0)[\bar{M}_u(0)u_{2,t}(0)]^2\right\}, \end{aligned}$$

where, suppressing the regime dependence,

$$\bar{M}_u = \frac{M_{12}\Sigma_{2|1}^{-1}}{M_{12}\Sigma_{2|1}^{-1}M_{21} + \Sigma_{11}^{-1}}, \quad \bar{V}_u = (M_{12}\Sigma_{2|1}^{-1}M_{21} + \Sigma_{11}^{-1})^{-1}.$$

Expressions (38), (40), and (39) can be plugged into (34) to evaluate the likelihood function.

4.2 Stochastic Volatility

To incorporate stochastic volatility into the analysis we interpret the VAR as a state-space model with latent volatility states $\zeta_t = [\zeta_1, \dots, \zeta_n]'$. Let $Y_{t_1:t_2}$ and $\zeta_{t_1:t_2}$ denote the sequences y_{t_1}, \dots, y_{t_2} and $\zeta_{t_1}, \dots, \zeta_{t_2}$ and θ be the vector of model parameters. The joint density of observations and states is given by

$$p(Y_{1:T}, \zeta_{1:T} | \theta) = \prod_{t=1}^T p(y_t | \zeta_t, Y_{1:t-1}, \theta) p(\zeta_t | \zeta_{t-1}, \theta). \quad (41)$$

The density $p(\zeta_t | \zeta_{t-1})$ is obtained from the laws of motion of the exogenously evolving volatilities in (18). The measurement equation, which depends on the current volatility state ζ_t and lags of the observables y_t is obtained from the vector autoregressive law of motion of y_t .

The likelihood function is obtained by integrating out the sequence $\zeta_{1:T}$ from the joint density of observables and states. We do so sequentially, using a particle filter that approximates:

$$p(Y_{1:T} | \theta) = \prod_{t=1}^T \int \int p(y_t | \zeta_t, Y_{1:t-1}, \theta) p(\zeta_t | \zeta_{t-1}, \theta) p(\zeta_{t-1} | Y_{1:t-1}, \theta) d\zeta_t d\zeta_{t-1}. \quad (42)$$

A particle filter represents the density $p(\zeta_t | Y_{1:t}, \theta)$ through a swarm of particles $\{\zeta_t^j, W_t^j\}_{j=1}^M$ with the property that posterior expectations $\mathbb{E}[h(\zeta_t) | Y_{1:t}, \theta]$ can be approximated by Monte Carlo averages of the form $\frac{1}{M} \sum_{j=1}^M h(\zeta_t^j) W_t^j$.

We use a simple bootstrap particle filter (BSPF) which was originally proposed by Gordon and Salmond (1993). In order to convert the time $t-1$ particle swarm $\{\zeta_{t-1}^j, W_{t-1}^j\}_{j=1}^M$ into a time t particle swarm $\{\zeta_t^j, W_t^j\}_{j=1}^M$ the BSPF simulates the law-of-motion of the states in (18) forward by one period for each ζ_{t-1}^j to obtain a $\tilde{\zeta}_t^j$. The time $t-1$ particle weights are updated based on $p(y_t | \tilde{s}_t^j, Y_{1:t-1}, \theta)$, which is identical to the likelihood increment in the homoskedastic version of the model, except that D needs to be replaced by $D_t(\tilde{\zeta}_t^j)$. To avoid a degeneracy of the particle weights, the particles can be resampled. As a by-product, the BSPF generates a stochastic approximation of the likelihood function that we denote by

$$\hat{p}(Y_{1:T} | \theta).$$

A detailed description of the algorithm based on the exposition in Herbst and Schorfheide (2015) is provided in the Online Appendix.

4.3 Identification

Based on data exclusively from regime $s = 1$ our SVAR is only partially identified. First, the $s = 1$ regime provides no information about Φ_{12}^{Δ} . Second, the number of free parameters in the matrices $A(1)$ and D exceed the number of estimable non-redundant reduced-form covariance parameters in $\Sigma(1)$. Bayesian inference remains valid, but there exist functions of model parameters for which the prior distribution is not updated; see, for instance, Poirier (1998) and Moon and Schorfheide (2012). In turn, a thoughtful specification of the prior distribution becomes important. The Baumeister and Hamilton (2018)-style prior is formulated based on parameterization that facilitates prior elicitation and imposes sign restrictions on some of the entries in the $A(1)$ matrix which reduces the size of the identified set.

Mavroeidis (2020) formally shows that the transition of the economy into the $s = 0$ regime generates additional identifying restrictions, which are akin to identification via heteroskedasticity and provides an identification analysis for his model specifications. It is fairly straightforward to demonstrate that our model satisfies a necessary condition for identification. Consider the following static example: $y_t' A(s) = \epsilon_t$, where $\epsilon_t \sim N(0, I)$ and our D matrix is absorbed into the definition of $A(s)$. From the $s = 1$ regime we can estimate $\mathbb{V}[y_{1,t}|y_{1,t} > 0]$, $\mathbb{E}[y_{2,t}|y_{1,t}]$, and $\mathbb{V}[y_{2,t}|y_{1,t}]$, which in a Gaussian framework with mean-zero shocks generates the usual $(n + 1)/2$ restrictions. From the $s = 0$ we can estimate $\mathbb{E}[y_{2,t}|y_{1,t} = 0]$ and $\mathbb{V}[y_{2,t}|y_{1,t} = 0]$, leading to an additional $(n - 1) + (n - 1)n/2$ restrictions. Whether or not these restrictions are sufficient for identification depends on the number of additional parameters that characterize the $s = 0$ distribution. In our model it is $n - 1$. In our application $n = 3$ and one element of $A(1)$ is set to zero. Thus, translated into the static framework, we have $(9 - 1) + 2 = 10$ unknowns and $6 + 5 = 11$ restrictions for identification, which implies that the necessary condition is satisfied.

5 Posterior Computations via SMC

Because of fairly complicated nonlinear parameter restrictions generated by the piecewise-linear structure of our SVAR, the posterior distribution of the parameters is non-standard.

We use a sequential Monte Carlo (SMC) sampler to generate draws from the posterior of θ . SMC techniques have emerged as an attractive alternative to MCMC methods because they can be easily parallelized and, properly tuned, may produce more accurate approximations of posterior distributions than MCMC algorithms.⁹

SMC combines features of classic importance sampling and modern MCMC techniques. The starting point is the creation of a sequence of intermediate or bridge distributions $\{\pi_n(\theta)\}_{n=0}^{N_\phi}$ that converge to the target posterior distribution, i.e., $\pi_{N_\phi}(\theta) = \pi(\theta)$. At any stage the (intermediate) posterior distribution $\pi_n(\theta)$ is represented by a swarm of particles $\{\theta_n^i, W_n^i\}_{i=1}^N$ in the sense that the Monte Carlo average

$$\bar{h}_{n,N} = \frac{1}{N} \sum_{i=1}^N W_n^i h(\theta_n^i) \xrightarrow{a.s.} \mathbb{E}_{\pi_n}[h(\theta_n)] \quad (43)$$

as $N \rightarrow \infty$, for each $n = 0, \dots, N_\phi$. We adopt the convention that the weights W_n^i are normalized to average to one. The bridge distributions are posterior distributions constructed from stage- n likelihood functions:

$$\pi_n(\theta) = \frac{p_n(Y|\theta)p(\theta)}{\int p_n(Y|\theta)p(\theta)d\theta} \quad (44)$$

with the convention that $p_0(Y|\theta) = 1$, i.e., the initial particles are drawn from the prior, and $p_{N_\phi}(Y|\theta) = p(Y|\theta)$. We use likelihood tempering of the form

$$p_n(Y|\theta) = [p(Y|\theta)]^{\phi_n}, \quad \phi_n = \left(\frac{n}{N_\phi}\right)^\ell. \quad (45)$$

The tuning parameter ℓ controls the shape of the tempering schedule.

The SMC algorithm proceeds iteratively from $n = 0$ to $n = N_\phi$. Starting from stage $n - 1$ particles $\{\theta_{n-1}^i, W_{n-1}^i\}_{i=1}^N$, each stage n of the algorithm targets the posterior π_n and consists of three steps: *correction*, that is, reweighting the stage $n - 1$ particles to reflect the density in iteration n ; *selection*, that is, eliminating a highly uneven distribution of particle weights (degeneracy) by resampling the particles; and *mutation*, that is, propagating the particles forward using a Markov transition kernel to adapt the particle values to the stage n bridge density. Our implementation of the SMC algorithm follows Herbst and Schorfheide (2015) and is described in detail in the Online Appendix.

⁹See Liu (2001), Cappé, Moulines, and Ryden (2005), and Herbst and Schorfheide (2015) for textbook treatments.

Likelihood values $p(Y|\theta)$ are required in the correction and mutation steps. For the model specification with stochastic volatility the exact likelihood function $p(Y|\theta)$ is replaced by the BSPF approximation $\hat{p}(Y|\theta)$ discussed in Section 4.2. In this case, the SMC algorithm turns into an SMC² algorithm, which has been studied in Chopin, Jacob, and Papaspiliopoulos (2013). Just like particle MCMC algorithms, the SMC² algorithm generates an approximation of the exact posterior $p(\theta|Y) \propto p(Y|\theta)p(\theta)$ provided the stochastic likelihood approximation $\hat{p}(Y|\theta)$ is unbiased, which it is for the BSPF.

6 Empirical Analysis

We now estimate a three-variable SVAR on U.S. quarterly data. $y_{1,t}$ is defined as the federal funds rate and $y_{2,t}$ comprises measures of output gap and inflation. The output gap is defined as log real GDP minus the log potential output series published by the Congressional Budget Office. We take inflation to be the year-over-year changes ($\ln P_t - \ln P_{t-4}$) in the personal consumption expenditure deflator. The three series are plotted in Figure 1. Because, unlike in our model, the effective federal funds rate was never exactly equal to zero when the economy reached the ELB, we set interest rates below 25 basis points (bp) equal to zero. The ELB episode is indicated by the gray band in the data plot.

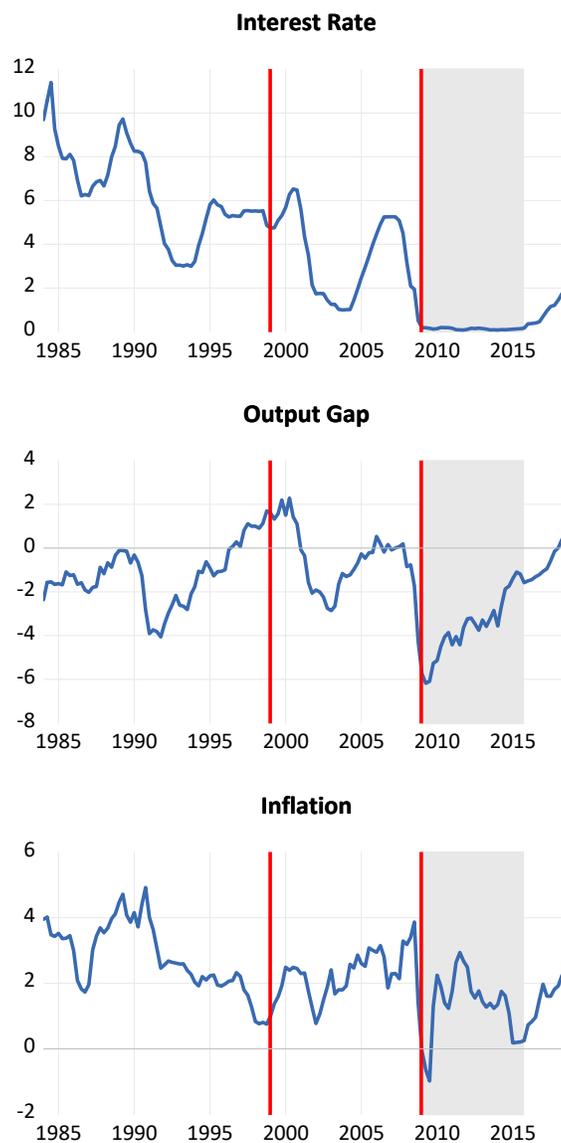
The subsequent empirical analysis is based on observations from 1984:Q1 to 2018:Q4. We use $p = 4$ lags for the SVAR. The SMC² algorithm described in Section 5 is used to generate draws θ^i , $i = 1, \dots, N$ from the posterior distribution.¹⁰ All the results presented subsequently are based on transformations of these draws. The parameter estimates are discussed in Section 6.1, the estimated impulse response functions are presented in Section 6.2, and the implied shadow rate is discussed in Section 6.3. All credible intervals reported in this section are 90% equal-tail-probability credible intervals.

6.1 Parameter and Volatility Estimates

Setting aside stochastic volatility in the structural innovations, there are two sources of nonlinearity in our SVAR. First, the interest rate is censored. This means, even if other variables depend on the interest rate linearly, their dynamics will change nonlinearly as the

¹⁰We use $N = 10,000$ particles for Algorithm 2 and $M = 100$ particles for the BSPF Algorithm 1. The SMC algorithm uses $N_\phi = 200$ stages with shape parameter $\ell = 2$. For the mutation we are using $N_{MH} = 3$ steps of a single-block MH algorithm.

Figure 1: Data

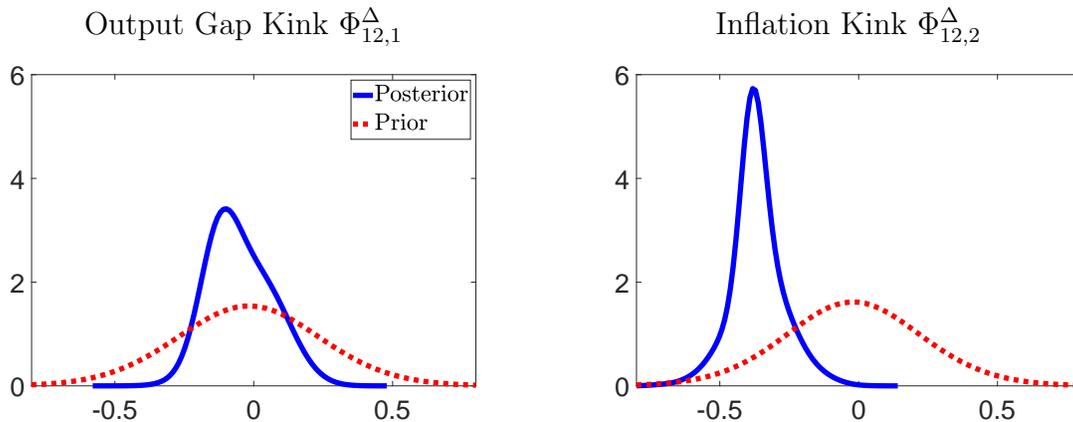


Notes: Grey bands indicate the ELB period from 2009:Q1 to 2015:Q4, during which the federal funds rate is below 25bp. Vertical lines denote dates for which we compute impulse response functions, which are 1999:Q1 and 2009:Q1.

interest rate reaches the ELB. Second, we allow for the laws of motion of the private-sector variables to depend on whether or not the interest rate is at or away from the ELB. This feature of our SVAR specification captures the potential nonlinearity of agents' decision rules in DSGE models with occasionally-binding constraints; see, for instance, ACHSV. The second nonlinearity is captured by a non-zero Φ_{12}^{Δ} .

Figure 2 shows prior and posterior densities for the two elements of Φ_{12}^{Δ} . The prior

Figure 2: Kinks in the Regression Functions



Notes: Marginal prior $p(\cdot)$ (dotted, red) and posterior $p(\cdot|Y)$ (solid, blue) densities. Panels depict kernel density estimates constructed from the output of the SMC algorithm.

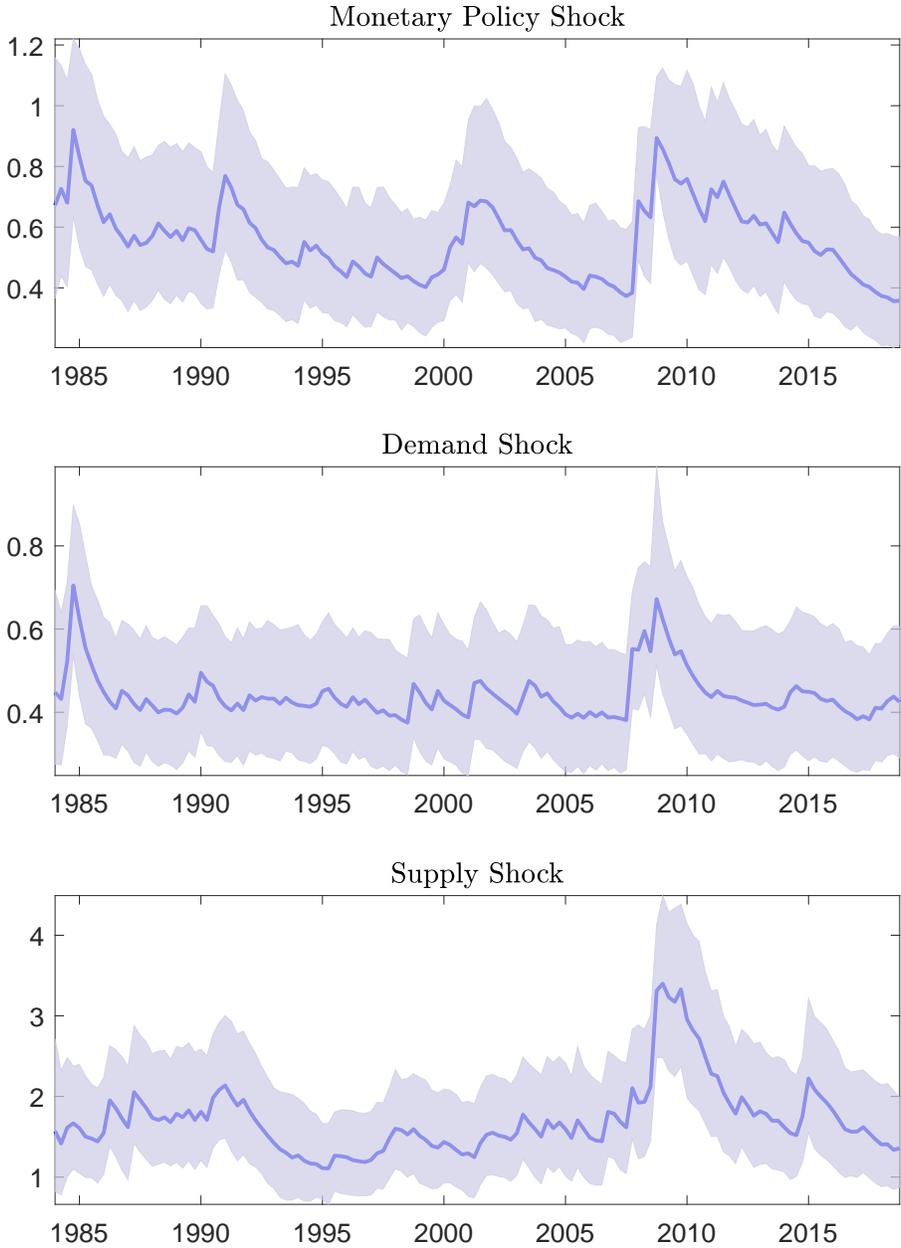
densities are almost symmetric around zero.¹¹ The posterior densities are clearly more concentrated than the prior densities. While the posterior for the output gap kink $\Phi_{12,2}^\Delta$ is centered around zero (the 90% credible interval ranges from -0.16 to 0.12), the posterior estimate of the inflation kink $\Phi_{12,1}$ is clearly negative. The 90% credible interval ranges from -0.52 to -0.21. These results suggest that inflation may respond to shocks differently at and away from the ELB, while output gap responses may be similar.

The third source of nonlinearity in our model comes from the stochastic volatility in the structural innovations. Figure 3 depicts the filtered volatilities for the three structural shocks, $\mathbb{E}[\sqrt{D_{ii,t}}|Y_{1:t}, \theta^i]$, computed using 10,000 draws θ^i from the posterior distribution, with $M = 1,000$ particles for each draw. In general, the stochastic volatility is more pronounced for the monetary policy shock and the supply shock than for the demand shock. During the Great Recession, volatility increases for all three shocks. The timing, however, differs. The demand shock volatility increases by more than 50% between 2007:Q3 to 2008:Q2 during the initial phase of the Great Recession and then peaks in 2008:Q4. The monetary policy shock volatility doubles in 2008:Q1 and then increases by another 40% in 2008:Q4. The volatility of the supply shock rises more gradually with a 56% jump in 2008:Q4 and peaks in 2009:Q4.

Finally, in Table 2 we compare the log marginal data densities (MDD) of various model specifications. Our preferred specification is the VAR(4) with stochastic volatility and non-zero Φ_{12}^Δ . Setting $\Phi_{12}^\Delta = 0$ reduces the log MDD by 8. The VAR(4) clearly dominates the

¹¹The adjustments described in Section 3.3 create a slight asymmetry.

Figure 3: Stochastic Volatility



Notes: Posterior distribution of filtered volatilities $\mathbb{E}[\sqrt{D_{ii,t}}|Y_{1:t}, \theta^i]$: mean (solid, blue) and 90% equal-tail-probability credible intervals (blue shades).

VAR(1) specifications. For both the VAR(1) and the VAR(4) the heteroskedastic version of the model is preferred to the homoskedastic specification.

Table 2: Marginal Data Densities

Specification	MDD
VAR(4), stochastic volatility (benchmark)	-303.34
VAR(4), stochastic volatility, $\Phi_{12}^{\Delta} = 0$	-311.34
VAR(4), homoskedastic	-322.92
VAR(1), stochastic volatility	-310.14
VAR(1), homoskedastic	-346.43

6.2 Responses to Shocks

To study the effects of interest-rate censoring and regime-specific coefficients and to document the parameter uncertainty, we consider three types of impulse responses. First, we fix the SVAR parameters at their posterior mean values and assume that the regime s persists forever. Second, we keep the parameters fixed at the posterior mean, but we allow the s_t regime and hence the coefficients of the $A\Phi$ representation to switch endogenously. Third, we generate bands that reflect parameter uncertainty.

In a nonlinear model like ours, impulse responses depend on lagged endogenous variables (and thus the date the shock hits), the size of the shock and, related, the level of volatility of the shocks for the duration of the impulse response. For all impulse responses, we condition on two specific dates: 1999:Q1 and 2009:Q1. These dates are indicated by the vertical lines in Figure 1. In 1999:Q1 output gap is positive, inflation is low, and the economy is far away from the ELB. In 2009:Q1 the economy is in the midst of the Great Recession with a large negative output gap, below-mean inflation, and interest rates at the ELB.

For all impulse responses we hold the level of volatility of the structural shocks constant throughout at the filtered value as measured on the date the initial shock hits the economy. Finally, we calibrate the size of the shocks as follows. First, we compute responses to minus-two-standard-deviation shocks for 1999:Q1. Then, for 2009:Q1, we scale the shocks as follows: the monetary policy shock is scaled so that the initial response of the shadow rate $y_{1,t}^*$ in 2009:Q1 is the same as the one for 1999:Q1. Second, demand and supply shocks are normalized such that the maximum absolute output responses are identical in 1999:Q1 and 2009:Q1.¹²

¹² This leads to the following scaling: all shocks are -2 standard deviations in $s = 1$ (1999:Q1) while in $s = 0$ (2009:Q1) we have -1.17 standard deviations for the monetary policy shock, -1.94 standard deviations for the demand shock and -0.99 standard deviations for the supply shock.

Fixed-Regime Impulse Responses. In Figure 4 we examine impulse responses that are computed under the assumption that the regime $s_t \in \{0, 1\}$ is fixed, where as we explained above we use 1999:Q1 and 2009:Q1 as representative dates for $s = 1$ and $s = 0$, respectively. This experiment can be interpreted as follows: the initial level $y_{1,t}^*$ is either so far above or below the censoring point $-\mu_1$, that the regime does not change for the next h periods in a forward simulation. We also fix the parameter vector θ at its posterior mean $\bar{\theta}$.

For the $s = 0$ regime, we generate impulse responses under two sets of regression functions: (i) the estimated regression functions and (ii) the $s = 1$ regression functions obtained by setting $\Phi_{12}^\Delta = 0$. In the graph, the former are labeled “ $s = 0$ ” whereas the latter are labeled “ $s = 0$ (linear).” We use the same scaling of shocks for these two sets of responses. A comparison between “ $s = 1$ ” and “ $s = 0$ (linear)” highlights the effect of the interest-rate censoring at the ELB, whereas a comparison between “ $s = 0$ (linear)” and $s = 0$ sheds light on the effect of the kink in the regression functions generated by the estimated Φ_{12}^Δ .

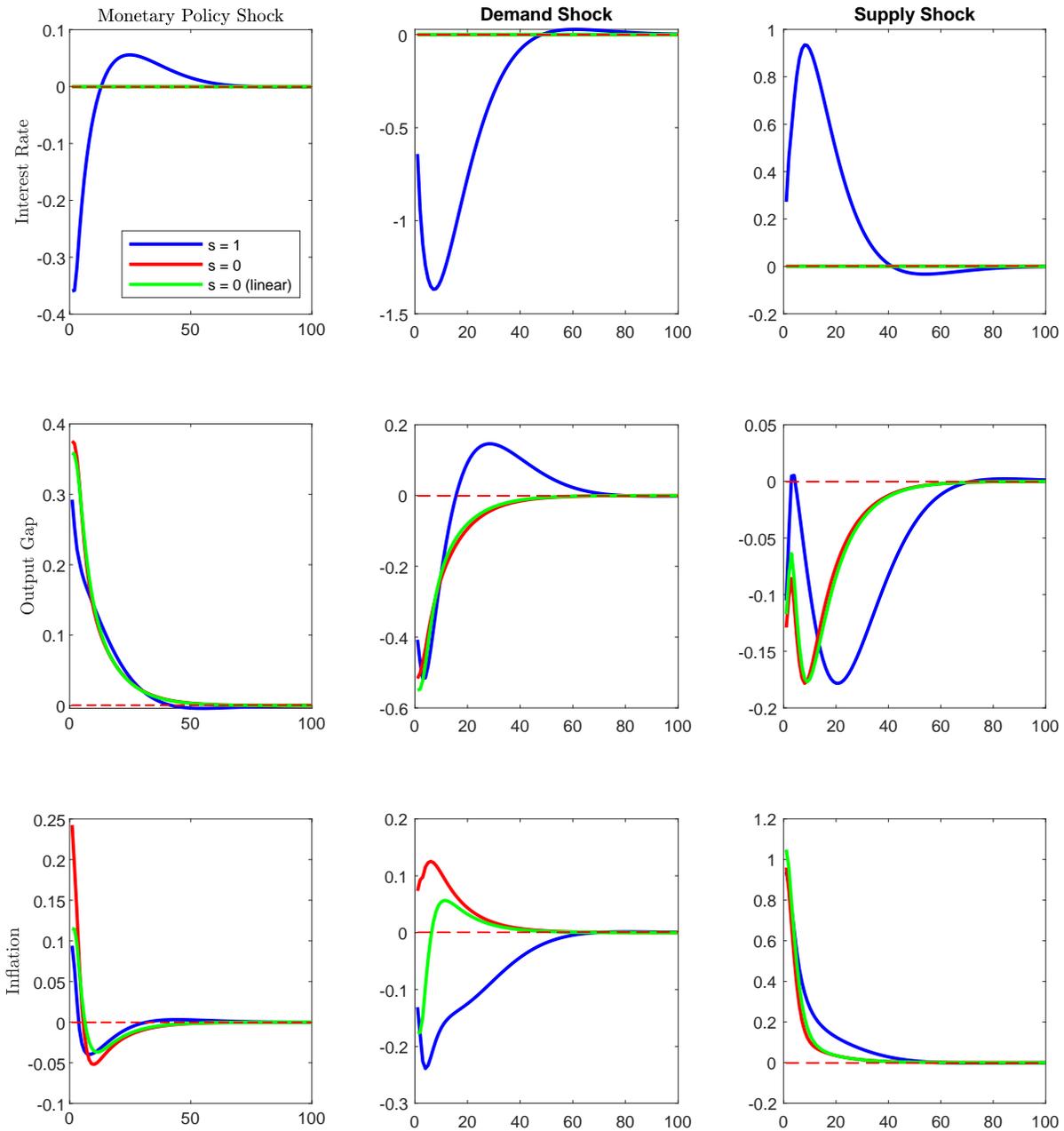
The blue lines in Figure 4 correspond to the $s = 1$ responses, and are computed from the implied estimates of $\Phi(1)$ and $\Phi^\epsilon(1)$. In response to a 36 basis point (bp) expansionary monetary policy shock the output gap rises by 29 bp and y-o-y inflation increases by 9 bp. The negative demand shock lowers output gap by 51 bp and inflation by 24 bp at its trough. In response to the fall in output and inflation, the central bank lowers nominal interest rates. The interest rate response is hump shaped and bottoms out at roughly -140 bp. The negative supply shock triggers a negative hump-shaped output-gap response with a trough at -18 bp and a 94 bp rise in inflation. The large increase in inflation leads the central bank to raise the nominal interest rate.

We now turn to the “ $s = 0$ (linear)” responses (green) that solely capture the effect of censoring at the ELB. By construction, the interest rate does not react to the shocks. Nonetheless, our model generates movements in output and inflation to the monetary policy shock, which can be interpreted as responses to changes in the shadow rate or the effects of unconventional monetary policy. Mechanically, these responses are generated by the vector $\Phi_{12}^\epsilon(s_t)$ in (9). This vector is obtained for $s = 0$ from the system of equations that link the $s = 0$ regression functions to the $s = 1$ regression functions; see (16).¹³

The $s = 0$ (linear) responses of the output gap and inflation to a monetary policy shock are somewhat larger than $s = 1$ responses, but very similar in shape to the non-ELB responses. Allowing for an estimated $\Phi_{12}^\Delta = 0$ essentially does not change the responses

¹³DSGE model solutions have the same feature: the monetary policy shock is a state variable and has an influence on decisions, regardless of whether or not the interest rate moves.

Figure 4: Responses to Negative Scaled Shocks with Fixed Regimes



Notes: The IRFs are computed by holding the regime s_t fixed. The blue lines represent the IRFs under the $s = 1$ regime (no censoring); the red lines are IRFs for the $s = 0$ regime (censoring and estimated value of Φ_{12}^Δ); the green lines are IRFs for the $s = 0$ regime with $\Phi_{12}^\Delta = 0$. The shocks are -2 standard deviation for the $s = 1$ regime and they are scaled for the $s = 0$ regime (see Footnote 12).

to a monetary policy shock. The responses to a supply shock are also very similar under the three scenarios. Only the inflation response to a demand shock exhibits substantial

differences across regimes. Under the $s = 1$ regime inflation falls drastically, reaching a trough of -24 bp, whereas under the $s = 0$ regime the inflation rises by 7 bp and continues to increase to 13 bp before it converges back to zero. $s = 0$ (linear) is an intermediate case in which inflation initially drops by 18 bp but then jumps to 5 bp before converging to zero in the long-run.¹⁴

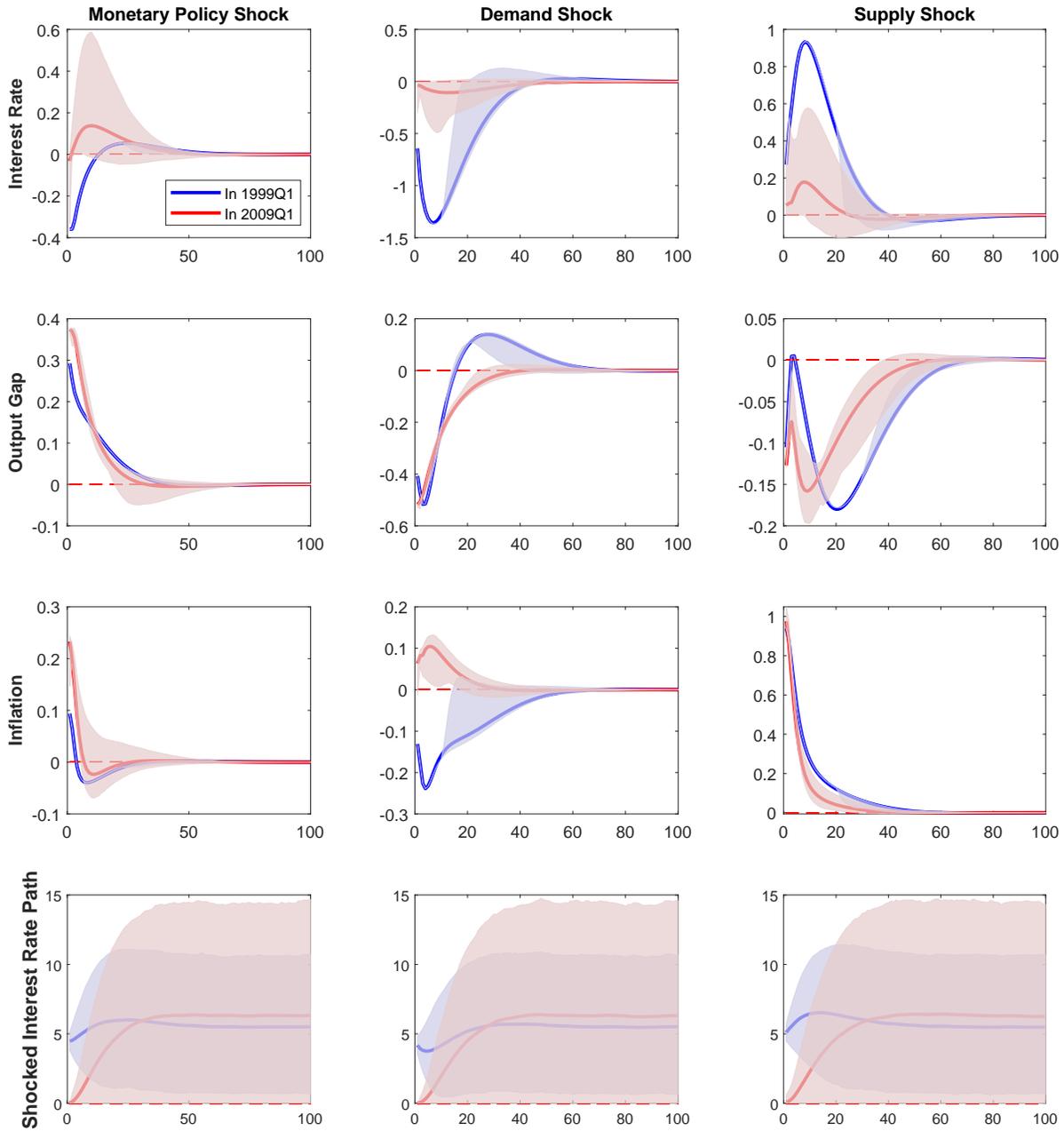
Ball and Mazumder (2011) and Hall (2011) pointed out that during the Great Recession inflation did not fall despite a large drop in output (gap), indicating a change in the positive relationship between the two variables implied by the Phillips curve. This observation was dubbed the missing deflation puzzle. Conditional on the posterior mean estimate $\bar{\theta}$ we computed estimates of the structural shocks (plotted in the Online Appendix). From 2008:Q4 to 2009:Q3 the demand shocks are estimated to be negative, depressing output. If the ELB regime parameters equaled their non-ELB regime counterparts, these shocks would have implied a substantial drop in inflation, just like a conventional Phillips curve, which did not happen. The model reconciles the empirical observation with a negative estimate of $\Phi_{12,2}^{\Delta}$ for inflation, which in turn delivers a rise in inflation in response to the demand shock in the ELB regime.

Impulse Responses With Regime Shifts. We continue to keep the parameter vector θ fixed at its posterior mean $\bar{\theta}$, but we now allow for the regime s_t to evolve over the impulse response as the economy goes in and out of the ELB. Doing so requires simulations because different realizations of the structural shocks may lead to different paths for the economy, including differences in when it is at or away from the ELB. Using $t+1$ to denote the date the shock hits the economy, starting from the initial condition x_{t+1} , we iterate the SVAR forward to obtain two different paths for the endogenous variables. Along the baseline trajectory, in simulation j , we draw all innovations $\epsilon_{i,t+h}^{j,0}$ from their respective $N(0, D_{i,t+1})$ distributions, not varying volatility with horizon h . We denote the resulting series $y_{t+1:t+H}^{j,0}$. To generate the shocked path $y_{t+1:t+H}^{j,1}$, we set $\epsilon_{1,t+1}^{j,1} = \epsilon_{1,t+1}^{j,0} + \delta$ and let $\epsilon_{i,t+h}^{j,1} = \epsilon_{i,t+h}^{j,0}$ for all other (i, h) . Here δ is the size of the shock as calibrated earlier. The impulse response is defined as the difference between the shocked and the baseline path. The simulation is repeated for $j = 1, \dots, M$ where we use $M = 10,000$.

In the top three rows of Figure 5 we plot the distribution of $y_{t+1:t+H}^{j,1} - y_{t+1:t+H}^{j,0}$ across j . The bands depict 90% credible intervals and the solid lines are means. In the absence of the nonlinearities triggered by the interest rate reaching the ELB, the bands collapse to

¹⁴The sign reversal of the inflation response relative to the $s = 1$ regime is also visible in the DSGE model based impulse responses reported in Figure 5 of Aruoba, Cuba-Borda, and Schorfheide (2018).

Figure 5: Responses in 1999:Q1 and 2009:Q1: Flexible Regimes and Shock Uncertainty



Notes: The top three panels show the distribution of the difference between a simulated baseline trajectory and a trajectory in which a scaled shock is subtracted from the simulated monetary policy / demand / supply shock in the initial period. See Footnote 12 for the scaling of the IRFs. The bottom panels show the level of the interest rate under the “shocked” trajectory. Bands represent 90% credible intervals and solid lines are means across $M = 10,000$ simulated paths.

a single line. The bottom panel of Figure 5 shows bands for the interest rate path along the “shocked” trajectories. Each time a trajectory j hits the ELB for the first time, the

difference between the shocked path and the baseline path will start to deviate from the fixed-regime $s = 1$ impulse response. Once a sufficiently large number of trajectories have hit the ELB, the response bands will start to fan out. Because of this cumulative effect, it is not necessary that the fanning out of the response bands coincides with the 5% quantiles of the interest rate distributions represented by the lower bounds of the shaded bands in the last row of Figure 5 reaching zero.

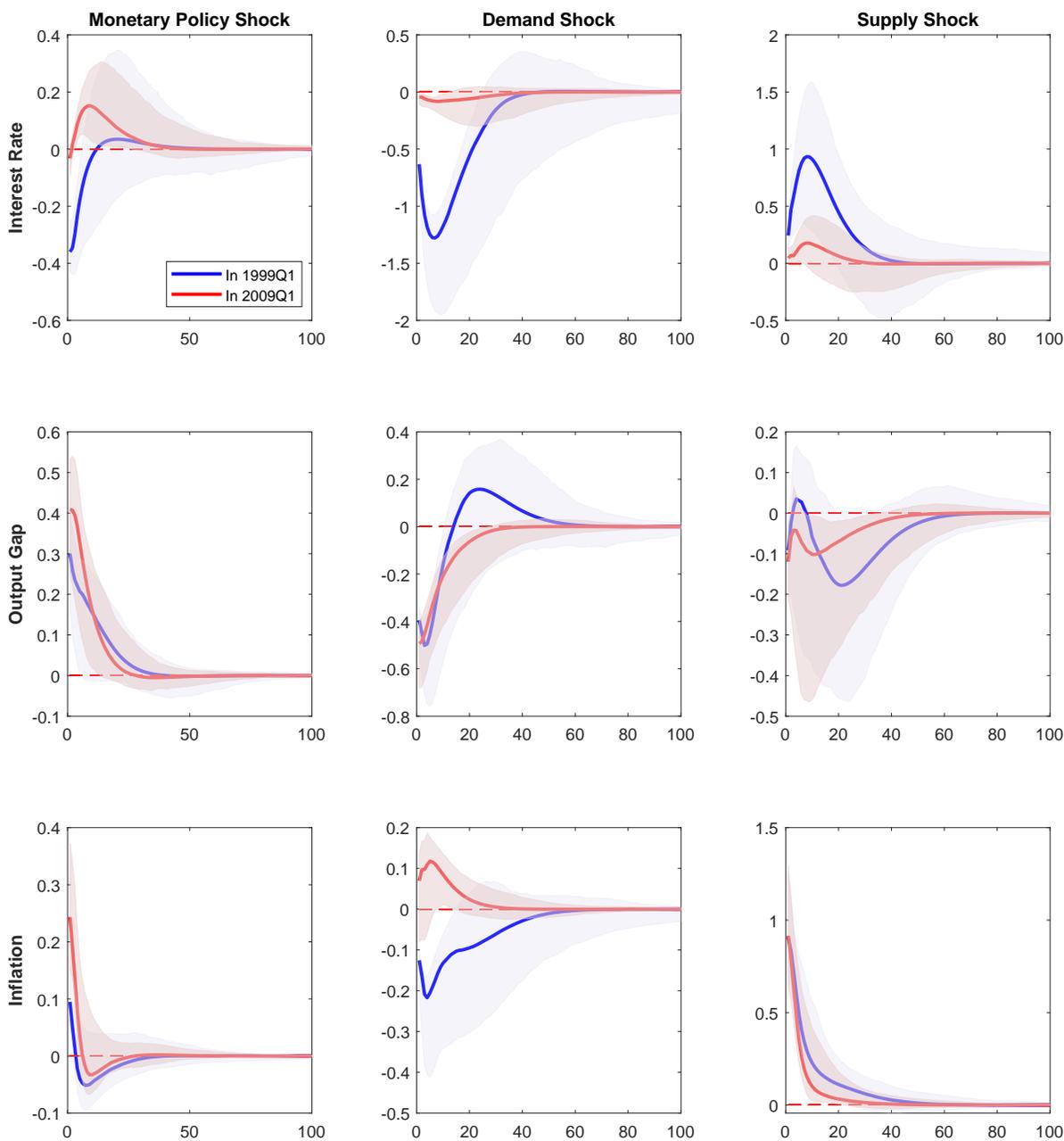
Qualitatively and quantitatively, the 1999:Q1 and 2009:Q1 responses look very similar to the $s = 1$ and $s = 0$ responses in Figure 4, respectively. All of the 1999:Q1 responses start out as single lines because initially the ELB is irrelevant. Some of the responses, in particular those to a demand shock, fan out after 25 to 50 periods. This means that there are a substantial number of trajectories along which interest rates are strictly positive ($s = 1$) under the baseline scenario, but hit the ELB ($s = 0$) at least once under the shocked scenario. Along these trajectories, the interest rate response reverts more quickly back to zero (because the interest rate drop is constrained by the ELB), and so does output gap and the inflation response. Visually, in Figure 5 the response along the far end (away from the average response) of the blue bands starts to look more similar to the mean response under the 2009:Q1 initial conditions.

The bands associated with the 2009:Q1 impulse responses fan out earlier and are typically wider than the ones for the 1999:Q1 responses because near the ELB the probability that the regimes along the shocked and unshocked trajectories differ is larger. Compared to the $s = 0$ responses in Figure 4 the 2009:Q1 IRFs are slightly tilted toward the $s = 1$ responses because under some of the simulated trajectories, the economy quickly moves away from the ELB. More generally, impulse responses for other dates will be somewhere in between the 1999:Q1 and 2009:Q1 responses, depending on how close the shadow rate is to the ELB.

Parameter Uncertainty. We will now explore the posterior uncertainty associated with the impulse response functions. The bands reported in Figure 6 represent 90% credible intervals that reflect posterior parameter uncertainty, in addition to the regime uncertainty we explored in Figure 5. The solid lines are pointwise medians of the impulse response posteriors. As in the previous figure, we compare responses based on the 1999:Q1 (blue) and 2009:Q1 (red) initial conditions. Figure 6 is generated by converting 10,000 draws θ^i from the posterior distribution into mean responses (each computed over 1,000 simulations) to the three structural shocks depicted as solid lines in Figure 5.¹⁵

¹⁵A comparison of prior and posterior impulse response bands is provided in the Online Appendix. The prior bands are substantially wider than the posterior bands indicating that the sample is informative about

Figure 6: Responses in 1999:Q1 and 2009:Q1: Flexible Regimes and Parameter Uncertainty



Notes: See Footnote 12 for the scaling of the IRFs. Bands represent 90% credible intervals and solid lines are pointwise posterior medians.

A comparison of the 1999:Q1 and 2009:Q1 bands sheds more light on the question whether the propagation of shocks is different at the ELB. In regard to interest rate responses, the

the propagation of shocks. Compared to the set-identified setting in BH, the ELB regime provides additional identifying information in our analysis.

answer is a trivial yes, because it is directly constrained by the ELB. More interesting is the comparison for output gap and inflation. Consistent with the insignificant estimate of $\Phi_{12,1}^{\Delta}$ for the output gap equation reported in Figure 2, we see that the output gap impulse response bands overlap for the two sets of initial conditions. Despite the overlap, the output gap response to the monetary policy shock appears to be slightly stronger when the economy is at the ELB. Figure 6 also reveals that there is a lot of parameter uncertainty regarding the output gap responses to a supply shock.

The inflation responses to a demand shock and a monetary policy shock, on the other hand, show some differences. This difference is most pronounced for the demand shock where the bands do not overlap for the first three years following the shock. While for 1999:Q1 the credible band for the inflation response is located below zero, for 2009:Q1 the inflation response is positive with high probability. As for the monetary policy shock, the unconventional expansionary monetary policy intervention at the ELB in 2009:Q1 is more inflationary on impact than away from the ELB in 1999:Q1, but the bands start to overlap in the subsequent periods.

Comparison to the Literature. To sum up our headline results, we find that (i) a negative monetary policy shock that creates the same size response in the shadow rate, is more inflationary at the ELB on impact and also generates a slightly larger response of the output gap. (ii) Negative demand shocks cause inflation to fall when the ELB is non-binding. At the ELB, on the other hand, the inflation response to a negative demand shock is positive. This difference persists for about three years. (iii) The output gap responses for demand and supply shocks do not significantly differ at and away from the ELB.

The papers by Ikeda, Li, Mavroeidis, and Zanetti (2020) and Johannsen and Mertens (2021) focus on differences in the effectiveness of monetary policy at and away from the ELB. Johannsen and Mertens (2021) find that monetary accommodation during the Great Recession would have provided more stimulus than at other times. Our results are qualitatively consistent with this finding: the output gap response is larger at the ELB relative to the response away from the ELB, though the credible sets overlap. Ikeda, Li, Mavroeidis, and Zanetti (2020) use a censored SVAR similar to the one in Mavroeidis (2020) and study the hypothesis of empirical irrelevance of the ELB using data from the U.S. and Japan. Their results show a similar inflation response to monetary policy on impact but a larger response at the ELB after two quarters and beyond, which is qualitatively similar to our results. Unlike us, they find a slight dampening instead of amplification of the output gap response. At the end, just as we do, they reject the irrelevance hypothesis.

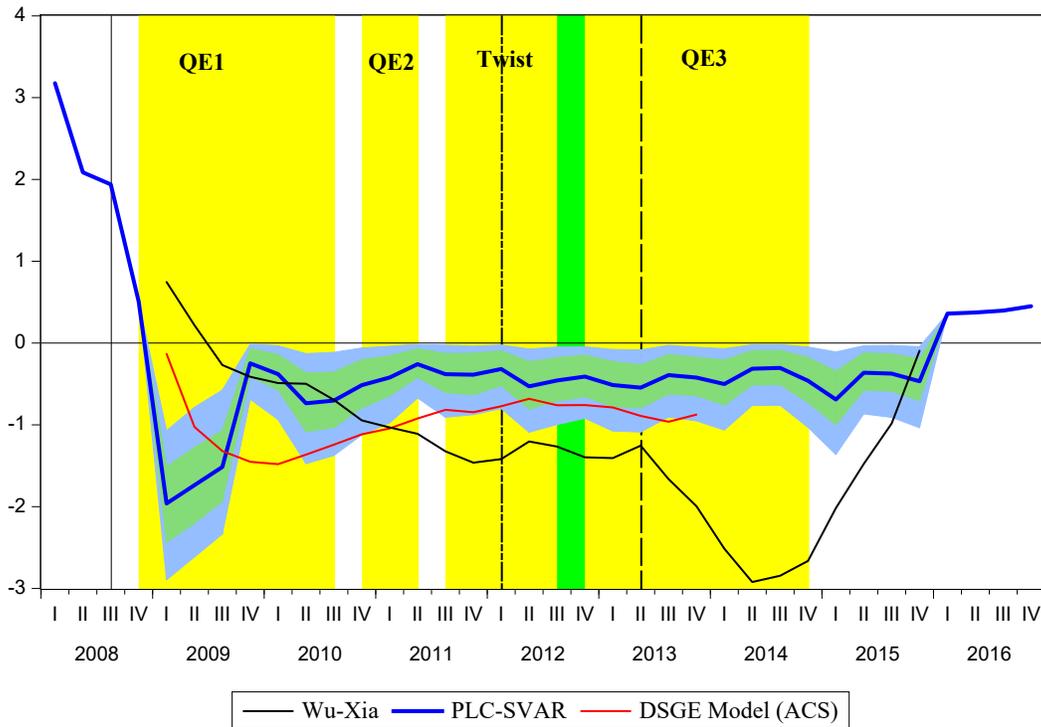
Our results (i) and (ii) are in contrast to the results from time-varying coefficient VARs presented in Debortoli, Galí, and Gambetti (2019). They do not find discernible differences between responses to a variety of shocks at and away from the ELB, which leads them to conclude that the ELB is empirically irrelevant. The densely parameterized and set-identified TVC-SVAR of Debortoli, Galí, and Gambetti (2019) makes it difficult to detect changes in the propagation of shocks based on a relatively short sample of ELB observations. We use more identifying information and only a small number of additional parameters to characterize the dynamics in the ELB regime. Thus, our approach is better equipped to detect differences in impulse response functions. All in all, we conclude that, as predicted by DSGE models with an ELB constraint, the ELB is not irrelevant for the propagation of shocks and this effect is measurable in a parsimonious SVAR framework that allows for changes in the private-sector behavior at the ELB.

6.3 Shadow Rate

Our model allows us to generate a shadow rate. It is given by an estimate of the latent variable $y_{1,t}^*$, which can then be adjusted to undo the effect of our re-centering: $\tilde{y}_{1,t}^* = y_{1,t}^* + \mu_1$. In the Online Appendix we provide a formula for $p(y_{1,t}^* | y_{1,t} = -\mu_1, y_{2,t}, x_t, \theta)$. Because in our SVAR x_t depends on $y_{1,t-1:t-p}$ instead of $y_{1,t-1:t-p}^*$ inference on the shadow rate is static and does not require dynamic filtering and smoothing. The posterior median estimate, 60% and 90% bands of the model-implied distribution of the shadow rate are plotted in Figure 7. The bands reflect uncertainty about the parameters θ and the shocks conditional on $(y_{1,t} = -\mu_1, y_{2,t}, x_t, \theta)$. For each parameter draw θ we condition on the mean (filtered) volatility estimate, which is allowed to vary over time. The inference exploits the correlation structure implied by $\Phi(0)$ and $\Sigma(0)$ between interest rates, on the one hand, and output gap and inflation on the other hand.

The estimated shadow rate drops to -2% in 2009 during the first large scale asset purchase intervention (QE1) of the Federal Reserve. From 2010 to 2015, during QE2, Operation Twist, and QE3 it hovered around -0.5%. This shows that the Federal Reserve was particularly aggressive in providing monetary stimulus in 2009 and reverted to a lesser but consistent stimulus for the rest of the ELB episode. It is also noteworthy that the slope of the shadow rate in early 2009 closely matches the slope in the federal funds rate just prior to the start of the ELB episode.

Figure 7: Shadow Rate



Notes: Blue line outside of the ELB episode: federal funds rate. During the ELB episode: 60% (green shade) and 90% (blue shade) credible intervals. Black line is Wu-Xia shadow rate and red line is the desired interest rate from the DSGE model of Aruoba, Cuba-Borda, and Schorfheide (2018). The solid vertical line denotes September 2008. The yellow shading indicates the quarters when the particular Fed program was active. During the green shaded area, both Operation Twist and QE3 were active. The dashed-dotted vertical line shows January 2012, when the formal inflation target was announced. The dashed vertical line shows the “taper tantrum” episode in June 2013.

Qualitatively, the time path of the shadow rate is consistent with the time path of the desired interest rate (red solid line) estimated with a small-scale New Keynesian DSGE in Aruoba, Cuba-Borda, and Schorfheide (2018). Quantitatively, the trough in the DSGE model implied shadow rate occurred about six months after the trough in the SVAR based shadow rate. Johannsen and Mertens (2021) produced a shadow rate based on their censored unobserved component model. Their shadow rate (not shown in the figure) is relatively flat at about -1% throughout the ELB episode and quite similar to our rate from 2010 onwards.

Finally, we provide a comparison with the yield-curve based shadow rate of Wu and Xia (2016). The Wu-Xia rate, rather than being based on a censored interest-rate feedback rule, is based on a censored affine term structure model and extracts information from yields on medium- and long-term bonds. Their shadow rate troughs in 2014, about five years

after the end of the Great Recession, about a year prior to the lift-off from the ELB. We find this somewhat implausible in view of the narrative evidence that the most significant interventions happened during and right after the Great Recession - which is consistent with our SVAR based estimates.¹⁶

7 Conclusion

We developed a structural VAR in which an occasionally-binding constraint generates censoring of one of the dependent variables. Once the censoring mechanism is triggered, we allow some of the coefficients for the remaining variables to change. By imposing that the regression functions are continuous at the censoring point, we can show that some mild parameter restrictions deliver a unique reduced form. The resulting model is more parsimonious than a time-varying-coefficient VAR and the switch in parameter values is tied to the censoring mechanism, which in our application is the ELB on nominal interest rates. An application to U.S. data provided evidence of a shift of parameters in the inflation equation which creates a more inflationary response to an expansionary monetary policy shock and a positive inflation response to a contractionary demand shock at the ELB.

References

- ARUOBA, S. B., P. CUBA-BORDA, K. HIGA-FLORES, F. SCHORFHEIDE, AND S. VILLALVAZO (2021): “Piecewise-Linear Approximations and Filtering for DSGE Models with Occasionally-Binding Constraints,” *Review of Economic Dynamics*, forthcoming.
- ARUOBA, S. B., P. CUBA-BORDA, AND F. SCHORFHEIDE (2018): “Macroeconomic Dynamics Near the ZLB: A Tale of Two Countries,” *Review of Economic Studies*, 85(1), 87–118.
- ARUOBA, S. B., AND F. SCHORFHEIDE (2011): “Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-Offs,” *American Economic Journal: Macroeconomics*, 3(1), 60–90.

¹⁶The shadow rate produced by Krippner (2015) looks even more extreme than the Wu-Xia rate, going down to about -6% in 2013.

- ASCARI, G., AND S. MAVROEIDIS (2020): “The Unbearable Lightness of Equilibria in a Low-Interest-Rate Environment,” *Manuscript, Oxford University*.
- BALL, L., AND S. MAZUMDER (2011): “Inflation Dynamics and the Great Recession,” *Brookings Papers on Economic Activity*, 42(1 (Spring)), 337–405.
- BAUMEISTER, C., AND J. D. HAMILTON (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.
- BLUNDELL, R., AND R. SMITH (1989): “Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models,” *Review of Economic Studies*, 56(1), 37–57.
- BOGNANNI, M., AND E. HERBST (2018): “A sequential Monte Carlo approach to inference in multiple-equation Markov-switching models,” *Journal of Applied Econometrics*, 33(1), 126–140.
- CAPPÉ, O., E. MOULINES, AND T. RYDEN (2005): *Inference in Hidden Markov Models*. Springer Verlag.
- CARRIERO, A., T. E. CLARK, M. MARCELLINO, AND E. MERTENS (2021): “Forecasting with Shadow-Rate VARs,” *Manuscript, Bundesbank*.
- CHOPIN, N., P. E. JACOB, AND O. PAPASPILIOPOULOS (2013): “SMC²: an efficient algorithm for sequential analysis of state space models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3), 397–426.
- DEBORTOLI, D., J. GALÍ, AND L. GAMBETTI (2019): “On the Empirical (Ir)relevance of the Zero Lower Bound Constraint,” in *NBER Macroeconomics Annual 2019*, ed. by M. Eichenbaum, E. Hurst, and J. Parker, vol. 34, pp. 141–170. University of Chicago Press.
- DEL NEGRO, M., AND F. SCHORFHEIDE (2009): “Monetary Policy with Potentially Misspecified Models,” *American Economic Review*, 99(4), 1415–1450.
- GIANNONE, D., M. LENZA, AND G. PRIMICERI (2015): “Prior Selection for Vector Autoregressions,” *Review of Economic and Statistics*, 97(2), 436–451.
- GORDON, N., AND D. SALMOND (1993): “A Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation,” *IEEE Proceedings-F*, 140, 107–113.

- GOURIEROUX, C., J. J. LAFFONT, AND A. MONFORT (1980): “Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes,” *Econometrica*, 48(3), 675–695.
- HALL, R. E. (2011): “The long slump,” *American Economic Review*, 101(2), 431–69.
- HERBST, E., AND F. SCHORFHEIDE (2014): “Sequential Monte Carlo Sampling for DSGE Models,” *Journal of Applied Econometrics*, 29(7), 1073–1098.
- (2015): *Bayesian Estimation of DSGE Models*. Princeton University Press.
- IKEDA, D., S. LI, S. MAVROEIDIS, AND F. ZANETTI (2020): “Testing the Effectiveness of Unconventional Monetary Policy in Japan and the United States,” *Manuscript, Oxford University*.
- IWATA, S., AND S. WU (2006): “Estimating Monetary Policy Effects When Interest Rates Are Close to Zero,” *Journal of Monetary Economics*, 53, 1395–1408.
- JOHANSEN, B. K., AND E. MERTENS (2021): “A Time-Series Model of Interest Rates with the Effective Lower Bound,” *Journal of Money, Credit, and Banking*, forthcoming.
- KRIPPNER, L. (2015): *Zero Lower Bound Term Structure Modeling: A Practitioner’s Guide*. Palgrave Macmillan.
- LIU, J. S. (2001): *Monte Carlo Strategies in Scientific Computing*. Springer Verlag.
- MAVROEIDIS, S. (2020): “Identification at the Zero Lower Bound,” *Manuscript, Oxford University*.
- MOON, H. R., AND F. SCHORFHEIDE (2012): “Bayesian and Frequentist Inference in Partially-Identified Models,” *Econometrica*, 80(2), 755–782.
- NELSON, F., AND L. OLSEN (1978): “Specification and Estimation of a Simultaneous-Equations Model with Limited Dependent Variables,” *International Economic Review*, 19(3), 695–709.
- POIRIER, D. (1998): “Revising Beliefs in Nonidentified Models,” *Econometric Theory*, 14(4), 483–509.
- WU, J. C., AND F. D. XIA (2016): “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound,” *Journal of Money, Credit and Banking*, 48(2-3), 253–291.

Online Appendix to “SVARs With Occasionally-Binding Constraints”

S. Boragan Aruoba, Marko Mlikota, Frank Schorfheide, and Sergio Villalvazo

This Appendix consists of the following sections:

- A. Proofs and Theoretical Derivations
- B. The Likelihood Function and Filtering
- C. Sequential Monte Carlo Algorithm
- D. Data Sources
- E. Additional Empirical Results

A Proofs and Theoretical Derivations

A.1 Proof of Proposition 1

Lemma 1 Consider a function $f(x) = \alpha_0 + \alpha_1 x$ and let

$$g(x) = f(x) + \gamma_0 + \gamma_1 x.$$

Suppose that $\alpha_1 \neq 0$ and either $\gamma_0 \neq 0$ or $\gamma_1 \neq 0$, then the set

$$\left\{ x \in \mathbb{R} \mid \text{sign}(g(x)) \neq \text{sign}(f(x)) \right\}$$

has non-zero Lebesgue measure.

Proof of Lemma 1. Suppose that $\alpha_1 > 0$. Then,

$$f(x) > 0 \implies \alpha_0 + \alpha_1 x > 0 \implies x > -\frac{\alpha_0}{\alpha_1}.$$

We now distinguish two cases. (i) Suppose that $(\alpha_1 + \gamma_1) > 0$. Then,

$$g(x) > 0 \implies (\alpha_0 + \gamma_0) + (\alpha_1 + \gamma_1)x > 0 \implies x > -\frac{\alpha_0 + \gamma_0}{\alpha_1 + \gamma_1}.$$

Thus,

$$\text{sign}(g(x)) \neq \text{sign}(f(x))$$

on the interval

$$\left(\min \left\{ -\frac{\alpha_0}{\alpha_1}, -\frac{\alpha_0 + \gamma_0}{\alpha_1 + \gamma_1} \right\}, \max \left\{ -\frac{\alpha_0}{\alpha_1}, -\frac{\alpha_0 + \gamma_0}{\alpha_1 + \gamma_1} \right\} \right). \quad (\text{A.1})$$

(ii) Suppose that $(\alpha_1 + \gamma_1) < 0$. Then,

$$g(x) > 0 \implies (\alpha_0 + \gamma_0) + (\alpha_1 + \gamma_1)x > 0 \implies x < -\frac{\alpha_0 + \gamma_0}{\alpha_1 + \gamma_1}.$$

Thus,

$$\text{sign}(g(x)) \neq \text{sign}(f(x))$$

on the set

$$\left(-\infty, \min \left\{ -\frac{\alpha_0}{\alpha_1}, -\frac{\alpha_0 + \gamma_0}{\alpha_1 + \gamma_1} \right\}\right) \cup \left(\max \left\{ -\frac{\alpha_0}{\alpha_1}, -\frac{\alpha_0 + \gamma_0}{\alpha_1 + \gamma_1} \right\}, \infty\right). \quad (\text{A.2})$$

The sets in (A.1) and (A.2) have non-zero Lebesgue measure. A similar calculation can be done for $a_1 < 0$ to complete the proof of the lemma. \square

We now proceed with the proof of Proposition 1.

Preliminaries. We will drop the time subscripts. Because of the linearity of the $s = 0$ and $s = 1$ regression functions, the functions $x'\Phi_{\cdot 1}(s = 0) + u_1(s = 0, \epsilon)$ and $x'\Phi_{\cdot 1}(s = 1) + u_1(s = 1, \epsilon)$ are continuous in (x, ϵ) . We regard x as predetermined. We begin with some algebraic manipulations. Recall that the shadow rate is equal to zero in the regimes $s = 0$ and $s = 1$ if, respectively,

$$x'\Phi_{\cdot 1}(0) + u_1(0, \epsilon) = -\mu_1 \quad \text{and} \quad x'\Phi_{\cdot 1}(1) + u_1(1, \epsilon) = -\mu_1. \quad (\text{A.3})$$

Thus, we define the distance from the threshold value $-\mu_1$ as

$$\begin{aligned} \Delta(s, x, \epsilon) &= x'\Phi_{\cdot 1}(s) + u_1(s, \epsilon) + \mu_1 & (\text{A.4}) \\ &= \frac{1}{A_{11}} \left[x_1(B_{11} - \Phi_{12}(s)A_{21}) + x_2'(B_{21} - \Phi_{22}(s)A_{21}) \right] \\ &\quad + \frac{1}{A_{11}} \left[\epsilon_1(1 - \Phi_{12}^\epsilon(s)A_{21}) - \epsilon_2'\Phi_{22}^\epsilon(s)A_{21} \right] + \mu_1 \\ &= \frac{1}{A_{11}} \Delta_x(s, x) + \frac{1}{A_{11}} \Delta_e(s, \epsilon) + \mu_1. \end{aligned}$$

We will begin by deriving a convenient relationship between $\Delta(1, x, \epsilon)$ and $\Delta(0, x, \epsilon)$. In order to capture potential discontinuities in the piecewise linear regression functions we introduce discrepancies Λ and let

$$\begin{aligned} \Phi_{j2}(0) &= \Phi_{j2}(1) + \frac{B_{j1} - \Phi_{j2}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{j2}, \quad j = 2, \dots, k-1 & (\text{A.5}) \\ \Phi_{k2}(0) &= \Phi_{k2}(1) + \frac{B_{k1} - \Phi_{k2}(1)A_{21} + \mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{k2} \\ \Phi_{12}^\epsilon(0) &= \Phi_{12}^\epsilon(1) + \frac{1 - \Phi_{12}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{12}^\epsilon \\ \Phi_{22}^\epsilon(0) &= \Phi_{22}^\epsilon(1) - \frac{\Phi_{22}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{22}^\epsilon. \end{aligned}$$

Using this notation, $\Delta_x(0, x)$ can be written as

$$\begin{aligned}
\Delta_x(0, x) &= x_1(B_{11} - \Phi_{12}(0)A_{21}) + x'_2(B_{21} - \Phi_{22}(0)A_{21}) \\
&= x_1(B_{11} - \Phi_{12}(0)A_{21}) \\
&\quad + \sum_{j=2}^{k-1} x_j \left(B_{j1} - \left[\Phi_{j2}(1) + \frac{B_{j,1} - \Phi_{j2}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{j2} \right] A_{21} \right) \\
&\quad + \left(B_{k1} - \left[\Phi_{k2}(1) + \frac{B_{k1} - \Phi_{k2}(1)A_{21} + \mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{k2} \right] A_{21} \right) \\
&= x_1(B_{11} - (\Phi_{12}(1) + \Phi_{12}^\Delta)A_{21}) + x'_2(B_{21} - \Phi_{22}(1)A_{21}) \\
&\quad - x'_2 \left[\frac{B_{21} - \Phi_{22}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{22} \right] A_{21} - \frac{\mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta A_{21},
\end{aligned} \tag{A.6}$$

where the second equality uses (A.5). The last equality, in slight abuse of notation, we let x'_2 be the row vector with elements x_j , B_{21} be the column vector with elements B_{j1} , Φ_{22} be the matrix with rows Φ_{j2} , and Λ_{22} be the matrix with rows Λ_{j2} for $j = 2, \dots, k$. Using the definition

$$\Delta_x(1, x) = x_1(B_{11} - \Phi_{12}(1)A_{21}) + x'_2(B_{21} - \Phi_{22}(1)A_{21})$$

we can rewrite (A.6) as follows:

$$\begin{aligned}
\Delta_x(0, x) &= \Delta_x(1, x) - x_1 \Phi_{12}^\Delta A_{21} - x'_2 \left[\frac{B_{21} - \Phi_{22}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{22} \right] A_{21} \\
&\quad - \frac{\mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta A_{21} \\
&= \Delta_x(1, x) - x_1 \frac{B_{11} - \Phi_{12}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta A_{21} - x'_2 \left[\frac{B_{21} - \Phi_{22}(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta + \Lambda_{22} \right] A_{21} \\
&\quad - \frac{\mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta A_{21} \\
&= \Delta_x(1, x) - \left(x_1(B_{11} - \Phi_{12}(1)A_{21}) + x'_2(B_{21} - \Phi_{22}(1)A_{21}) \right) \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \\
&\quad - x'_2 \Lambda_{22} A_{21} - \frac{\mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta A_{21} \\
&= \Delta_x(1, x) \left(1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right) - x'_2 \Lambda_{22} A_{21} - \frac{\mu_1 A_{11}}{B_{11} - \Phi_{12}(1)A_{21}} \Phi_{12}^\Delta A_{21}.
\end{aligned} \tag{A.7}$$

Similarly, we can derive a relationship between $\Delta_e(0, \epsilon)$ and $\Delta_e(1, \epsilon)$:

$$\begin{aligned}
\Delta_e(0, \epsilon) &\equiv \epsilon_1(1 - \Phi_{12}^\epsilon(0)A_{21}) - \epsilon'_2\Phi_{22}^\epsilon(0)A_{21} & (A.8) \\
&= \epsilon_1 \left(1 - \left(\Phi_{12}^\epsilon(1) + \frac{1 - \Phi_{12}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}}\Phi_{12}^\Delta + \Lambda_{12}^\epsilon \right) A_{21} \right) \\
&\quad - \epsilon'_2 \left(\Phi_{22}^\epsilon(1) - \frac{\Phi_{22}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}}\Phi_{12}^\Delta + \Lambda_{22}^\epsilon \right) A_{21} \\
&= \epsilon_1(1 - \Phi_{12}^\epsilon(1)A_{21}) - \epsilon'_2\Phi_{22}^\epsilon(1)A_{21} \\
&\quad - \epsilon_1 \frac{1 - \Phi_{12}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}}\Phi_{12}^\Delta A_{21} + \epsilon'_2 \frac{\Phi_{22}^\epsilon(1)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}}\Phi_{12}^\Delta A_{21} - (\epsilon_1\Lambda_{12}^\epsilon + \epsilon'_2\Lambda_{22}^\epsilon) A_{21} \\
&= \Delta_e(1, \epsilon) - \left(\epsilon_1(1 - \Phi_{12}^\epsilon(1)A_{21}) - \epsilon'_2\Phi_{22}^\epsilon(1)A_{21} \right) \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \\
&\quad - (\epsilon_1\Lambda_{12}^\epsilon + \epsilon'_2\Lambda_{22}^\epsilon) A_{21} \\
&= \Delta_e(1, \epsilon) \left(1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right) - (\epsilon_1\Lambda_{12}^\epsilon + \epsilon'_2\Lambda_{22}^\epsilon) A_{21}.
\end{aligned}$$

Combining (A.7) and (A.8) we get

$$\begin{aligned}
\Delta(0, x, \epsilon) &= \frac{1}{A_{11}}(\Delta_x(1, x) + \Delta_e(1, \epsilon)) \left(1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right) - \frac{\mu_1}{B_{11} - \Phi_{12}(1)A_{21}}\Phi_{12}^\Delta A_{21} \\
&\quad - \frac{1}{A_{11}}(x'_2\Lambda_{22} + \epsilon_1\Lambda_{12}^\epsilon + \epsilon'_2\Lambda_{22}^\epsilon) A_{21} + \mu_1 & (A.9) \\
&= \left(\frac{1}{A_{11}}\Delta_x(1, x) + \frac{1}{A_{11}}\Delta_e(1, \epsilon) + \mu_1 \right) \left(1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right) \\
&\quad - \frac{1}{A_{11}}(x'_2\Lambda_{22} + \epsilon_1\Lambda_{12}^\epsilon + \epsilon'_2\Lambda_{22}^\epsilon) A_{21}.
\end{aligned}$$

Because $\Delta(1, x, \epsilon)$ is defined as

$$\Delta(1, x, \epsilon) = \frac{1}{A_{11}}\Delta_x(1, x) + \frac{1}{A_{11}}\Delta_e(1, \epsilon) + \mu_1$$

we obtain

$$\Delta(0, x, \epsilon) = \Delta(1, x, \epsilon) \left[1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right] - \frac{1}{A_{11}}(x'_2\Lambda_{22} + \epsilon_1\Lambda_{12}^\epsilon + \epsilon'_2\Lambda_{22}^\epsilon) A_{21}. \quad (A.10)$$

Proof of Part (i). A necessary condition for uniqueness given (x, ϵ) is

$$\begin{aligned}
x'\Phi_{\cdot 1}(1) + u_1(1, \epsilon) + \mu_1 > 0 &\text{ implies } x'\Phi_{\cdot 1}(0) + u_1(0, \epsilon) + \mu_1 > 0 \\
x'\Phi_{\cdot 1}(0) + u_1(0, \epsilon) + \mu_1 < 0 &\text{ implies } x'\Phi_{\cdot 1}(1) + u_1(1, \epsilon) + \mu_1 < 0.
\end{aligned}$$

We will assume that $A_{11} > 0$ (otherwise inequalities will be reversed). We can write the condition more compactly as

$$\text{sign}(\Delta(1, x, \epsilon)) = \text{sign}(\Delta(0, x, \epsilon)). \quad (\text{A.11})$$

Let $x = \tilde{x}$, partition $\epsilon = [\epsilon_j, \epsilon_{(-j)}]$, and let $\epsilon_{(-j)} = \tilde{\epsilon}_{(-j)}$. Notice that $\Delta(s, \tilde{x}, [\epsilon_j, \tilde{\epsilon}_{(-j)}])$ is a linear function of ϵ_j . Thus, we can define $\alpha_0(\cdot)$ and α_1 such that

$$f(\epsilon_j) = \left[1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right] \Delta(1, \tilde{x}, [\epsilon_j, \tilde{\epsilon}_{(-j)}]) = \alpha_0(\tilde{x}, \tilde{\epsilon}_{(-j)}) + \alpha_1 \epsilon_j.$$

Moreover, we can define $\gamma_0(\cdot)$ and γ_1 such that

$$-\frac{1}{A_{11}} \left(\tilde{x}'_2 \Lambda_{22} + \epsilon_1([\epsilon_j, \tilde{\epsilon}_{(-j)}]) \Lambda_{12}^\epsilon + \epsilon'_2([\epsilon_j, \tilde{\epsilon}_{(-j)}]) \Lambda_{22}^\epsilon \right) A_{21} = \gamma_0((\tilde{x}, \tilde{\epsilon}_{(-j)})) + \gamma_1 \epsilon_j.$$

We can now choose j such that $\alpha_1 \neq 0$ and a pair $(\tilde{x}, \tilde{\epsilon}_{(-j)})$ such that either $\gamma_0 \neq 0$ or $\gamma_1 \neq 0$. Part (i) of the proposition now follows from Lemma 1. \square

Proof of Part (ii). Given x , let $\mathcal{E}^0(x)$ denote the set of innovations for which the ELB becomes binding. Thus, for any $\epsilon \in \mathcal{E}^0(x)$

$$x' \Phi_{\cdot 1}(0) + u_1(0, \epsilon) = -\mu_1 \quad \text{and} \quad x' \Phi_{\cdot 1}(1) + u_1(1, \epsilon) = -\mu_1,$$

which implies that

$$\Delta(0, x, \epsilon) = \Delta(1, x, \epsilon) = 0.$$

Now consider $\epsilon \notin \mathcal{E}^0(x)$. Setting the Λ terms in (A.10) to zero, we obtain

$$\Delta(0, x, \epsilon) = \Delta(1, x, \epsilon) \left[1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} \right]. \quad (\text{A.12})$$

Note that (A.11) is satisfied for every (x, ϵ) if and only if

$$1 - \frac{\Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} > 0. \quad (\text{A.13})$$

Using the definition $\Phi_{12}^\Delta = \Phi_{12}^0 - \Phi_{12}^1$ we can manipulate this condition as follows:

$$\begin{aligned} & \frac{B_{11} - \Phi_{12}(1)A_{21} - \Phi_{12}^\Delta A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} > 0 \\ \implies & \frac{B_{11} - \Phi_{12}(0)A_{21}}{B_{11} - \Phi_{12}(1)A_{21}} > 0 \\ \implies & \text{sign}(B_{11} - \Phi_{12}(0)A_{21}) = \text{sign}(B_{11} - \Phi_{12}(1)A_{21}), \end{aligned} \quad (\text{A.14})$$

which yields the expressions in the statement in Part (ii). \square

A.2 Proof of Proposition 2

We drop the t subscripts throughout.

Proposition 3 (Comparison of $\mathcal{M}(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s))$ and $\mathcal{M}(A_{12}(s))$)

$$(i) \quad \mathcal{M}(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s)) \supset \mathcal{M}(A_{12}(s))$$

$$(ii) \quad \mathcal{M}_{PLC}(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s)) = \mathcal{M}(A_{12}(s))$$

Proof of Part (i). We combine the A -representation of the monetary policy rule (7) and the A -representation of the private sector equations (17) to derive the Φ representation of the private sector equations and compare it with (9). Solve (7) for y_1^*

$$y_1^* = \frac{1}{A_{11}}(x'B_{.1} + \epsilon_1 - y_2'A_{21}). \quad (\text{A.15})$$

Combining (17) and (A.15) yields

$$\frac{1}{A_{11}}(x'B_{.1} + \epsilon_1 - y_2'A_{21})A_{12}(s) + y_2'A_{22} = x'B_{.2}(s) + \epsilon_2'.$$

Re-arranging terms yields

$$y_2' \left(A_{22} - \frac{A_{21}A_{12}(s)}{A_{11}} \right) = x' \left(B_{.2}(s) - B_{.1} \frac{A_{12}(s)}{A_{11}} \right) - \epsilon_1 \frac{A_{12}(s)}{A_{11}} + \epsilon_2'.$$

Finally, we can solve for y'_2 to obtain

$$y'_2 = x' \left(B_{.2}(s) - B_{.1} \frac{A_{12}(s)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(s)}{A_{11}} \right)^{-1} \quad (\text{A.16})$$

$$- \epsilon_1 \frac{A_{12}(s)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(s)}{A_{11}} \right)^{-1} + \epsilon'_2 \left(A_{22} - \frac{A_{21}A_{12}(s)}{A_{11}} \right)^{-1}.$$

This generates, in our notation, the matrices

$$\Phi_{.2}(s), \quad \Phi_{12}^\epsilon(s), \quad \Phi_{22}^\epsilon(s). \quad (\text{A.17})$$

Statement (i) of the proposition follows by noting that $\mathcal{M}(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s))$ has $(n-1) \times (k+n)$ free Φ coefficients in regimes $s = 0$ and $s = 1$. $\mathcal{M}(A_{12}(s))$ has $(n-1) \times k + n^2$ free Φ coefficients in the $s = 1$ regime, but only $n - 1$ additional unrestricted coefficients in the $s = 0$ regime. \square

Proof of Part (ii). We will show that $\mathcal{M}_{PLC}(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s)) \supseteq \mathcal{M}(A_{12}(s))$ and $\mathcal{M}_{PLC}(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s)) \subseteq \mathcal{M}(A_{12}(s))$.

Step 1: $\mathcal{M}_{PLC}(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s)) \supseteq \mathcal{M}(A_{12}(s))$. We show that the $(\Phi_{.2}(s), \Phi_{.2}^\epsilon(s))$ regression functions implied by $\mathcal{M}(A_{12}(s))$ specifications are continuous at points in the state-space in which the economy hits the ELB. We deduce from (7) and (17) that interest rates hit the ELB when

$$-\mu_1 A_{11} + y'_2 A_{21} = x' B_{.1} + \epsilon_1$$

$$-\mu_1 A_{12}(1) + y'_2 A_{22} = x' B_{.2}(1) + \epsilon'_2.$$

The second equation implies that

$$y'_2 = (\mu_1 A_{12}(1) + x' B_{.2}(1) + \epsilon'_2) A_{22}^{-1}.$$

Thus, we obtain that the ELB is reached when

$$-\mu_1 A_{11} + (\mu_1 A_{12}(1) + x' B_{.2}(1) + \epsilon'_2) A_{22}^{-1} A_{21} = x' B_{.1} + \epsilon_1,$$

or

$$\begin{aligned}\epsilon_1 &= -\mu_1 A_{11} + (\mu_1 A_{12}(1) + x' B_{\cdot 2}(1) + \epsilon'_2) A_{22}^{-1} A_{21} - x' B_{\cdot 1}. \\ &= \mu_1 (-A_{11} + A_{12}(1) A_{22}^{-1} A_{21}) + \epsilon'_2 A_{22}^{-1} A_{21} + x' (-B_{\cdot 1} + B_{\cdot 2}(1) A_{22}^{-1} A_{21}).\end{aligned}\tag{A.18}$$

Note that unlike in the main paper, here we solved for ϵ_1 instead of x_1 .

Using the $\Phi(s)$ matrices implied by the $A(s)$ representation from (A.16), the continuity requirement for the $\Phi(s)$ regression functions can be stated as

$$\begin{aligned}0 &= x' \left(B_{\cdot 2}(0) - B_{\cdot 1} \frac{A_{12}(0)}{A_{11}} \right) \left(A_{22} - \frac{A_{21} A_{12}(0)}{A_{11}} \right)^{-1} \\ &\quad - \epsilon_1 \frac{A_{12}(0)}{A_{11}} \left(A_{22} - \frac{A_{21} A_{12}(0)}{A_{11}} \right)^{-1} + \epsilon'_2 \left(A_{22} - \frac{A_{21} A_{12}(0)}{A_{11}} \right)^{-1} \\ &\quad - x' \left(B_{\cdot 2}(1) - B_{\cdot 1} \frac{A_{12}(1)}{A_{11}} \right) \left(A_{22} - \frac{A_{21} A_{12}(1)}{A_{11}} \right)^{-1} \\ &\quad + \epsilon_1 \frac{A_{12}(1)}{A_{11}} \left(A_{22} - \frac{A_{21} A_{12}(1)}{A_{11}} \right)^{-1} - \epsilon'_2 \left(A_{22} - \frac{A_{21} A_{12}(1)}{A_{11}} \right)^{-1},\end{aligned}\tag{A.19}$$

for all $(x, \epsilon_1, \epsilon_2)$ such that (A.18) is satisfied.

After plugging (A.18) into (A.19) to eliminate ϵ_1 , the coefficients on ϵ'_2 are

$$\begin{aligned}&\left(A_{22} - \frac{A_{21} A_{12}(0)}{A_{11}} \right)^{-1} - \left(A_{22} - \frac{A_{21} A_{12}(1)}{A_{11}} \right)^{-1} \\ &- A_{22}^{-1} A_{21} \frac{A_{12}(0)}{A_{11}} \left(A_{22} - \frac{A_{21} A_{12}(0)}{A_{11}} \right)^{-1} + A_{22}^{-1} A_{21} \frac{A_{12}(1)}{A_{11}} \left(A_{22} - \frac{A_{21} A_{12}(1)}{A_{11}} \right)^{-1} \\ &= A_{22}^{-1} - A_{22}^{-1} = 0.\end{aligned}\tag{A.20}$$

Recall that

$$B_{\cdot 2}(s) = [B_{12} | \dots | B_{(k-1)2} | B_{k2}(s)], \quad B_{k2}(s) = \begin{cases} 0 & \text{if } s = 1 \\ \mu_1 (A_{12}(1) - A_{12}(0)) & \text{if } s = 0 \end{cases}.$$

Using the fact that B_{j2} , $j = 1, \dots, k-1$, is not state dependent, the coefficients on x_j ,

$j = 1, \dots, k - 1$ are

$$\begin{aligned}
& \left(B_{j2} - B_{j1} \frac{A_{12}(0)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} - \left(B_{j2} - B_{j1} \frac{A_{12}(1)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
& - (-B_{j1} + B_{j2}A_{22}^{-1}A_{21}) \frac{A_{12}(0)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& + (-B_{j1} + B_{j2}A_{22}^{-1}A_{21}) \frac{A_{12}(1)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \tag{A.21} \\
= & \left(B_{j2} - B_{j1} \frac{A_{12}(0)}{A_{11}} - (-B_{j1} + B_{j2}A_{22}^{-1}A_{21}) \frac{A_{12}(0)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& - \left(B_{j2} - B_{j1} \frac{A_{12}(1)}{A_{11}} + (-B_{j1} + B_{j2}A_{22}^{-1}A_{21}) \frac{A_{12}(1)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
= & \left(B_{j2} - B_{j2}A_{22}^{-1}A_{21} \frac{A_{12}(0)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& - \left(B_{j2} + B_{j2}A_{22}^{-1}A_{21} \frac{A_{12}(1)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
= & B_{j2}A_{22}^{-1} - B_{j2}A_{22}^{-1} = 0.
\end{aligned}$$

We now turn to $x_k = 1$ and will use that $B_{k2}(1) = 0$ and $B_{k2}(0) = \mu_1(A_{12}(1) - A_{12}(0))$.

$$\begin{aligned}
& \left(\mu_1(A_{12}(1) - A_{12}(0)) - B_{k1} \frac{A_{12}(0)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& - \left(0 - B_{k1} \frac{A_{12}(1)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
& - (-B_{k1} + 0) \frac{A_{12}(0)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} + (-B_{k1} + 0) \frac{A_{12}(1)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
& - \mu_1(-A_{11} + A_{12}(1)A_{22}^{-1}A_{21}) \frac{A_{12}(0)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& + \mu_1(-A_{11} + A_{12}(1)A_{22}^{-1}A_{21}) \frac{A_{12}(1)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
= & \mu_1(A_{12}(1) - A_{12}(0)) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& + \mu_1(A_{11} - A_{12}(1)A_{22}^{-1}A_{21}) \frac{A_{12}(0)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& - \mu_1(A_{11} - A_{12}(1)A_{22}^{-1}A_{21}) \frac{A_{12}(1)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
= & \mu_1(A_{12}(1) - A_{12}(0)) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& + \mu_1 \left(A_{12}(0) - A_{12}(1)A_{22}^{-1}A_{21} \right) \frac{A_{21}A_{12}(0)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& - \mu_1 \left(A_{12}(1) - A_{12}(1)A_{22}^{-1}A_{21} \right) \frac{A_{21}A_{12}(1)}{A_{11}} \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
= & \mu_1 \left(A_{12}(1)A_{22}^{-1}A_{22} - A_{12}(1)A_{22}^{-1} \frac{A_{21}A_{12}(0)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& - \mu_1 \left(A_{12}(1)A_{22}^{-1}A_{22} - A_{12}(1)A_{22}^{-1} \frac{A_{21}A_{12}(1)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
= & \mu_1 A_{12}(1)A_{22}^{-1} \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(0)}{A_{11}} \right)^{-1} \\
& - \mu_1 A_{12}(1)A_{22}^{-1} \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right) \left(A_{22} - \frac{A_{21}A_{12}(1)}{A_{11}} \right)^{-1} \\
= & 0.
\end{aligned}$$

From (A.20), (A.21), and (A.22) we deduce that (A.19) holds for all $(x, \epsilon_1, \epsilon_2)$ such that (A.18) is satisfied. \square

Step 2: $\mathcal{M}_{PLC}(\Phi_{\cdot 2}(s), \Phi_{\cdot 2}^\epsilon(s)) \subseteq \mathcal{M}(A_{12}(s))$. Starting from a specification in $\mathcal{M}_{PLC}(\Phi_{\cdot 2}(s), \Phi_{\cdot 2}^\epsilon(s))$

we derive the A representation and show that A_{22} and B_{j2} , $j = 1, \dots, k - 1$ are constant across states. Solving (7) for $\epsilon_{1,t}$ and substituting the resulting expression into (9) yields:

$$y'_2 = x' \Phi_{.2}(s) + (y_1^* A_{11} + y'_2 A_{21} - x' B_{.1}) \Phi_{12}^\epsilon(s) + \epsilon'_2 \Phi_{22}^\epsilon(s).$$

Re-arranging terms yields

$$\begin{aligned} & -y_1^* A_{11} \Phi_{12}^\epsilon(s) [\Phi_{22}^\epsilon(s)]^{-1} + y'_2 (I - A_{21} \Phi_{12}^\epsilon(s)) [\Phi_{22}^\epsilon(s)]^{-1} \\ & = x' (\Phi_{.2}(s) - B_{.1} \Phi_{12}^\epsilon(s)) [\Phi_{22}^\epsilon(s)]^{-1} + \epsilon'_2. \end{aligned}$$

From a comparison to (17) we obtain:

$$\begin{aligned} A_{12}(s) &= -A_{11} \Phi_{12}^\epsilon(s) [\Phi_{22}^\epsilon(s)]^{-1} \\ A_{22}(s) &= (I - A_{21} \Phi_{12}^\epsilon(s)) [\Phi_{22}^\epsilon(s)]^{-1} \\ B_{.2}(s) &= (\Phi_{.2}(s) - B_{.1} \Phi_{12}^\epsilon(s)) [\Phi_{22}^\epsilon(s)]^{-1}. \end{aligned} \tag{A.22}$$

We now have to verify that $A_{22}(1) = A_{22}(0)$ and that $B_{.2}(1) = B_{j2}(0)$, $j = 1, \dots, k - 1$, $B_{k2}(1) = 0$, and $B_{k2}(0) = \mu_1(A_{12}(1) - A_{12}(0))$.

Consider the continuity restriction. The ELB becomes binding when

$$-\mu_1 A_{11} + y'_2 A_{21} = x' B_{.1} + \epsilon_1.$$

Plugging this expression into (9) yields

$$y'_2 = x' \Phi_{.2}(s) + (-\mu_1 A_{11} + y'_2 A_{21} - x' B_{.1}) \Phi_{12}^\epsilon(s) + \epsilon'_2 \Phi_{22}^\epsilon(s).$$

Solving for y_2 yields

$$y'_2 = \left(x' (\Phi_{.2}(s) - B_{.1} \Phi_{12}^\epsilon(s)) - \mu_1 A_{11} \Phi_{12}^\epsilon(s) + \epsilon'_2 \Phi_{22}^\epsilon(s) \right) (I - A_{21} \Phi_{12}^\epsilon(s))^{-1}.$$

Continuity requires that y_2 is the same regardless of ϵ_2 and x for $s = 0$ and $s = 1$, that is:

$$\begin{aligned}
\epsilon_2 \text{ term : } & \Phi_{22}^\epsilon(1)(I - A_{21}\Phi_{12}^\epsilon(1))^{-1} & (A.23) \\
& = \Phi_{22}^\epsilon(0)(I - A_{21}\Phi_{12}^\epsilon(0))^{-1} \\
x_j \text{ term : } & (\Phi_{j2}(1) - B_{j1}\Phi_{12}^\epsilon(1))(I - A_{21}\Phi_{12}^\epsilon(1))^{-1} \\
& = (\Phi_{j2}(0) - B_{j1}\Phi_{12}^\epsilon(0))(I - A_{21}\Phi_{12}^\epsilon(0))^{-1}, \quad j = 1, \dots, k-1 \\
x_k \text{ term : } & (\Phi_{k2}(1) - B_{k1}\Phi_{12}^\epsilon(1) - \mu_1 A_{11}\Phi_{12}^\epsilon(1))(I - A_{21}\Phi_{12}^\epsilon(1))^{-1} \\
& = (\Phi_{k2}(0) - B_{k1}\Phi_{12}^\epsilon(0) - \mu_1 A_{11}\Phi_{12}^\epsilon(0))(I - A_{21}\Phi_{12}^\epsilon(0))^{-1}.
\end{aligned}$$

Re-arranging the first equality in display (A.23) yields

$$(I - A_{21}\Phi_{12}^\epsilon(0))[\Phi_{22}^\epsilon(0)]^{-1} = (I - A_{21}\Phi_{12}^\epsilon(1))[\Phi_{22}^\epsilon(1)]^{-1}. \quad (A.24)$$

and therefore implies that

$$A_{22}(0) = A_{22}(1). \quad (A.25)$$

We can rewrite (A.24) as

$$(I - A_{21}\Phi_{12}^\epsilon(1))^{-1}(I - A_{21}\Phi_{12}^\epsilon(0)) = [\Phi_{22}^\epsilon(1)]^{-1}\Phi_{22}^\epsilon(0). \quad (A.26)$$

Combining (A.26) with the second set of equations in display (A.23), we obtain

$$(\Phi_{j2}(1) - B_{j1}\Phi_{12}^\epsilon(1))[\Phi_{22}^\epsilon(1)]^{-1}\Phi_{22}^\epsilon(0) = (\Phi_{j2}(0) - B_{j1}\Phi_{12}^\epsilon(0)), \quad j = 1, \dots, k-1.$$

Thus, we can deduce that

$$B_{j2}(1) = B_{j2}(0), \quad j = 1, \dots, k-1. \quad (A.27)$$

as required.

Finally, we will manipulate the third equation in display (A.23) as follows. Following the same steps as for the x_j term, we can write

$$\begin{aligned}
& (\Phi_{k2}(1) - B_{k1}\Phi_{12}^\epsilon(1) - \mu_1 A_{11}\Phi_{12}^\epsilon(1))[\Phi_{22}^\epsilon(1)]^{-1} \\
& = (\Phi_{k2}(0) - B_{k1}\Phi_{12}^\epsilon(0) - \mu_1 A_{11}\Phi_{12}^\epsilon(0))[\Phi_{22}^\epsilon(0)]^{-1}.
\end{aligned}$$

Using the expressions for $A_{12}(s)$ and $B_{k2}(s)$ in (A.22) we obtain

$$B_{k2}(1) + \mu_1 A_{12}(1) = B_{k2}(0) + \mu_1 (A_{12}(0)).$$

Moreover, because the intercepts $\Phi_{k2}(1)$ and B_{k1} are equal to zero, we can deduce that

$$B_{k2}(1) = 0, \quad B_{k2}(0) = \mu_1 (A_{12}(1) - A_{12}(0)). \quad \square$$

A.3 An Alternative Formulation of the Necessary and Sufficient Condition in Proposition 1

Solving (13) for $\epsilon_{1,t}$ yields

$$\epsilon_{1,t} = -x'_t \frac{(B_{.1} - \Phi_{.2}(1)A_{21})}{1 - \Phi_{12}^\epsilon(1)A_{21}} + \epsilon'_{2,t} \frac{\Phi_{22}^\epsilon(1)A_{21}}{1 - \Phi_{12}^\epsilon(1)A_{21}} - \mu_1 \frac{A_{11}}{1 - \Phi_{12}^\epsilon(1)A_{21}}. \quad (\text{A.28})$$

For continuity, we require that for all $(x, \epsilon_1, \epsilon_2)$ such that (A.28) is satisfied the following equality holds:

$$x' \Phi_{.2}(0) + \epsilon_1 \Phi_{12}^\epsilon(0) + \epsilon'_2 \Phi_{22}^\epsilon(0) = x' \Phi_{.2}(1) + \epsilon_1 \Phi_{12}^\epsilon(1) + \epsilon'_2 \Phi_{22}^\epsilon(1) \quad (\text{A.29})$$

After replacing ϵ_1 by (A.28), we deduce from the coefficients on ϵ_2 that

$$\Phi_{22}^\epsilon(0) = \Phi_{22}^\epsilon(1) - \frac{\Phi_{22}^\epsilon(1)A_{21}}{1 - \Phi_{12}^\epsilon(1)A_{21}} \Phi_{12}^{\epsilon\Delta} \quad (\text{A.30})$$

where

$$\Phi_{12}^{\epsilon\Delta} = \Phi_{12}^\epsilon(0) - \Phi_{12}^\epsilon(1).$$

From the coefficients on x' we can deduce the following. Starting with $j = 1, \dots, k-1$

$$\Phi_{j2}(0) = \Phi_{j2}(1) + \frac{(B_{j1} - \Phi_{j2}(1)A_{21})}{1 - \Phi_{12}^\epsilon(1)A_{21}} \Phi_{12}^{\epsilon\Delta}. \quad (\text{A.31})$$

For $j = k$ we obtain

$$\Phi_{k2}(0) = \Phi_{k2}(1) + \frac{(B_{k1} - \Phi_{k2}(1)A_{21})}{1 - \Phi_{12}^\epsilon(1)A_{21}} \Phi_{12}^{\epsilon\Delta} + \mu_1 \frac{A_{11}}{1 - \Phi_{12}^\epsilon(1)A_{21}} \Phi_{12}^{\epsilon\Delta}. \quad (\text{A.32})$$

We can verify that the restrictions in (A.30) to (A.32) are equivalent to (16). Consider

the condition in (A.31) for $j = 1$:

$$\frac{\Phi_{12}^{\epsilon\Delta}}{1 - \Phi_{12}^{\epsilon}(1)A_{21}} = \frac{\Phi_{12}^{\Delta}}{B_{11} - \Phi_{12}^{\Delta}(1)A_{21}}. \quad (\text{A.33})$$

Plugging (A.33) into (A.30) to (A.32) reproduces (16). Moreover, using (A.33), we can rewrite the condition in Proposition 1(ii) as

$$\text{sign}(1 - \Phi_{12}^{\epsilon}(1)A_{21}) = \text{sign}(1 - \Phi_{12}^{\epsilon}(0)A_{21})$$

which is equivalent to

$$\kappa = \frac{A_{11} - A_{11}\Phi_{12}^{\epsilon}(0)A_{21}}{A_{11} - A_{11}\Phi_{12}^{\epsilon}(1)A_{21}} > 0. \quad (\text{A.34})$$

Using (A.16) we can write

$$A_{11} - A_{11}\Phi_{12}^{\epsilon}(s)A_{21} = A_{11} + A_{12}(s) (A_{22} - A_{21}A_{11}^{-1}A_{12}(s))^{-1} A_{21}. \quad (\text{A.35})$$

From the formulas for the inverse of a partitioned matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}$$

we obtain the relationship

$$(D - CA^{-1}B)^{-1} = D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}.$$

Thus, we can rewrite (A.35) as follows

$$\begin{aligned} & A_{11} - A_{11}\Phi_{12}^{\epsilon}(s)A_{21} && (\text{A.36}) \\ &= A_{11} + A_{12}(s) (A_{22} - A_{21}A_{11}^{-1}A_{12}(s))^{-1} A_{21} \\ &= A_{11} + A_{12}(s)A_{22}^{-1}A_{21} + A_{12}(s)A_{22}^{-1}A_{21}(A_{11} - A_{12}(s)A_{22}^{-1}A_{21})^{-1} A_{12}(s)A_{22}^{-1}A_{21} \\ &= A_{11} + \lambda(s) + \frac{\lambda^2(s)}{A_{11} - \lambda(s)} \\ &= \frac{1}{A_{11} - \lambda(s)} (A_{11}^2 - A_{11}\lambda(s) + \lambda(s)A_{11} - \lambda^2(s) + \lambda^2(s)) \\ &= \frac{A_{11}^2}{A_{11} - \lambda(s)}, \end{aligned}$$

where $\lambda(s) = A_{12}(s)A_{22}^{-1}A_{21}$ is scalar. In turn,

$$\kappa = \frac{A_{11} - A_{12}(1)A_{22}^{-1}A_{21}}{A_{11} - A_{12}(0)A_{22}^{-1}A_{21}}, \quad (\text{A.37})$$

which is identical to the condition provided in Proposition 1 of Mavroeidis (2020) because in his notation the role of 1 and 2 subscripts are exchanged. \square

B The Likelihood Function and Filtering

B.1 Derivations for the Likelihood Function

The derivations for $p(y_{1,t}|\cdot)$ in (38) and $p(y_{2,t}|y_{1,t} > -\mu_1, \cdot)$ in (40) are straightforward. We will focus on the derivation of $p(y_{2,t}|y_{1,t} = -\mu_1, \cdot)$ in (39). In the latter case, the specific value of $u_{1,t}(0)$ is unknown. Write

$$p(y_{2,t}|y_{1,t} = -\mu_1, \cdot) = \int p(y_{2,t}|u_{1,t})p(u_{1,t}|y_{1,t} = -\mu_1, \cdot)du_{1,t}. \quad (\text{A.38})$$

We begin by deriving the density of $u_{1,t}$ conditional on $y_{1,t} = -\mu_1$:

$$p(u_{1,t}|y_{1,t} = -\mu_1, \cdot) = \frac{p_N(u_{1,t}; 0, \Sigma_{11}(0))\mathbb{I}\{u_{1,t}(0) \leq -x'_t\Phi_{\cdot 1}(0) - \mu_1\}}{F_N\left(-\frac{x'_t\Phi_{\cdot 1}(1) + \mu_1}{\sqrt{\Sigma_{11}(1)}}\right)}. \quad (\text{A.39})$$

Combining (A.38) and (A.39) yields:

$$\begin{aligned} p(y_{2,t}|y_{1,t} = -\mu_1, \cdot) &= \left[F_N\left(-\frac{x'_t\Phi_{\cdot 1}(1) + \mu_1}{\sqrt{\Sigma_{11}(1)}}\right) \right]^{-1} \int p_N(u_{2,t}(0); u_{1,t}M_{12}(0), \Sigma_{2|1}(0)) \\ &\quad \times p_N(u_{1,t}; 0, \Sigma_{11}(0))\mathbb{I}\{u_{1,t}(0) \leq -x'_t\Phi_{\cdot 1}(0) - \mu_1\}du_{1,t}. \end{aligned} \quad (\text{A.40})$$

In the subsequent steps we will evaluate the integral.

To simplify the notation we drop the $s_t = 0$ argument from Σ and Φ matrices. The key

terms that appear in the integral in (A.40) can be manipulated as follows:

$$\begin{aligned}
& (u'_{2,t} - u_{1,t}M_{12})\Sigma_{2|1}^{-1}(u_{2,t} - M_{21}u_{1,t}) + u_{1,t}\Sigma_{11}^{-1}u_{1,t} \\
&= u'_{2,t}\Sigma_{2|1}^{-1}u_{2,t} - 2u_{1,t}M_{12}\Sigma_{2|1}^{-1}u_{2,t} + u_{1,t}M_{12}\Sigma_{2|1}^{-1}M_{21}u_{1,t} + u_{1,t}\Sigma_{11}^{-1}u_{1,t} \\
&= u'_{2,t}\Sigma_{2|1}^{-1}u_{2,t} + (M_{12}\Sigma_{2|1}^{-1}M_{21} + \Sigma_{11}^{-1}) \left(u_{1,t} - \frac{M_{12}\Sigma_{2|1}^{-1}}{M_{12}\Sigma_{2|1}^{-1}M_{21} + \Sigma_{11}^{-1}}u_{2,t} \right)^2 \\
&\quad - \frac{(M_{12}\Sigma_{2|1}^{-1}u_{2,t})^2}{M_{12}\Sigma_{2|1}^{-1}M_{21} + \Sigma_{11}^{-1}}.
\end{aligned}$$

Define

$$\bar{M}_u = \frac{M_{12}\Sigma_{2|1}^{-1}}{M_{12}\Sigma_{2|1}^{-1}M_{21} + \Sigma_{11}^{-1}}, \quad \bar{V}_u = (M_{12}\Sigma_{2|1}^{-1}M_{21} + \Sigma_{11}^{-1})^{-1}$$

Notice that

$$\begin{aligned}
(2\pi)^{-1/2}|\bar{V}_u|^{-1/2} &\times \int \exp \left\{ -\frac{1}{2\bar{V}_u}(u_{1,t} - \bar{M}_u u_{2,t})^2 \right\} \mathbb{I} \left\{ \frac{u_{1,t} - \bar{M}_u u_{2,t}}{\sqrt{\bar{V}_u}} \leq -\frac{x'_t\Phi_{\cdot 1} + \mu_1 + \bar{M}_u u_{2,t}}{\sqrt{\bar{V}_u}} \right\} du_{1,t} \\
&= F_N \left(-\frac{x'_t\Phi_{\cdot 1} + \mu_1 + \bar{M}_u u_{2,t}}{\sqrt{\bar{V}_u}} \right).
\end{aligned}$$

Therefore, we obtain Equation (40) in the main text:

$$\begin{aligned}
& p(y_{2,t}|y_{1,t} = -\mu_1, \cdot) \\
&= (2\pi)^{-(n-1)/2}|\Sigma_{2|1}(0)|^{-1/2}|\Sigma_{11}(0)|^{-1/2}|\bar{V}_u(0)|^{1/2} \left[F_N \left(-\frac{x'_t\Phi_{\cdot 1}(1) + \mu_1}{\sqrt{\Sigma_{11}(1)}} \right) \right]^{-1} \\
&\quad \times F_N \left(-\frac{x'_t\Phi_{\cdot 1}(0) + \mu_1 + \bar{M}_u(0)u_{2,t}(0)}{\sqrt{\bar{V}_u(0)}} \right) \\
&\quad \times \exp \left\{ -\frac{1}{2}u'_{2,t}(0)\Sigma_{2|1}^{-1}(0)u_{2,t}(0) + \frac{1}{2}\bar{V}_u^{-1}(0)[\bar{M}_u(0)u_{2,t}(0)]^2 \right\} \quad \square
\end{aligned}$$

B.2 Particle Filter

We use a bootstrap particle filter (BSPF) to approximate the likelihood function in the model with stochastic volatility. In the description of the filter we denote the latent state by s_t .

Algorithm 1 (Bootstrap Particle Filter)

1. **Initialization.** Draw the initial particles from the distribution $s_0^j \stackrel{iid}{\sim} p(s_0|\theta)$ and set $W_0^j = 1$, $j = 1, \dots, M$.

2. **Recursion.** For $t = 1, \dots, T$:

(a) **Forecasting s_t .** Draw \tilde{s}_t^j from the state-transition density $p(\tilde{s}_t|s_{t-1}^j, \theta)$.

(b) **Forecasting y_t .** Define the incremental weights

$$\tilde{w}_t^j = p(y_t|\tilde{s}_t^j, Y_{1:t-1}, \theta) \quad (\text{A.41})$$

The predictive density $p(y_t|Y_{1:t-1}, \theta)$ can be approximated by

$$\hat{p}(y_t|Y_{1:t-1}, \theta) = \frac{1}{M} \sum_{j=1}^M \tilde{w}_t^j W_{t-1}^j. \quad (\text{A.42})$$

(c) Define the normalized weights

$$\tilde{W}_t^j = \tilde{w}_t^j W_{t-1}^j / \frac{1}{M} \sum_{j=1}^M \tilde{w}_t^j W_{t-1}^j. \quad (\text{A.43})$$

(d) **Selection.** Resample the particles, for instance, via multinomial resampling. Let $\{s_t^j\}_{j=1}^M$ denote M iid draws from a multinomial distribution characterized by support points and weights $\{\tilde{s}_t^j, \tilde{W}_t^j\}$ and set $W_t^j = 1$ for $j = 1, \dots, M$. An approximation of $\mathbb{E}[h(s_t)|Y_{1:t}, \theta]$ is given by $\bar{h}_{t,M} = \frac{1}{M} \sum_{j=1}^M h(s_t^j) W_t^j$.

3. **Likelihood Approximation.** The approximation of the log-likelihood function is given by

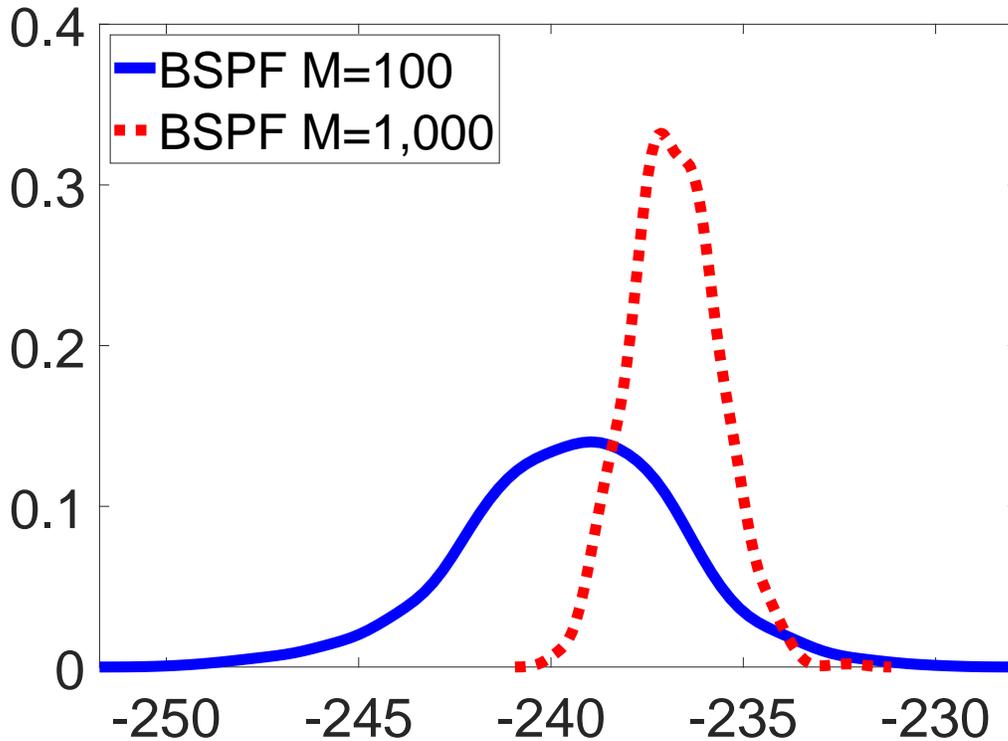
$$\ln \hat{p}(Y_{1:T}|\theta) = \sum_{t=1}^T \ln \left(\frac{1}{M} \sum_{j=1}^M \tilde{w}_t^j W_{t-1}^j \right). \quad (\text{A.44})$$

Figure A-1 shows the accuracy of the BSPF. Conditional on the posterior mean $\bar{\theta}$ we run the BSPF $N_{run} = 100$ times. Because it is a stochastic algorithm, the likelihood approximations differ across runs. The figure depicts kernel density estimates of the distribution of log likelihood values across runs. The more concentrated the density, the more accurate the approximation. The accuracy increases with the number of particles M . The SMC² algorithm implicitly averages out the approximation errors.

B.3 Derivations for the Shadow Rate

Homoskedasticity. The shadow rate can be characterized through the density $p(y_{1,t}^*|y_{1,t} = -\mu_1, y_{2,t}, x_t, \theta)$. Recall that our model conditional on $y_{1,t} = -\mu_1$ (and, hence, $s_t = 0$) has

Figure A-1: Accuracy of Particle Filter Approximation of Log Likelihood



Notes: Density plots are based on $N_{run} = 100$ runs of the BSPF ($M = 100$ is solid blue and $M = 1,000$ is dotted red). The parameter value θ is fixed at the posterior mean $\hat{\theta}$.

the reduced-form representation

$$y_{1,t}^* = x_t' \Phi_{\cdot 1}(0) + u_{1,t}(0), \quad y_{2,t}' = x_t' \Phi_{\cdot 2}(0) + u_{2,t}'(0). \quad (\text{A.45})$$

Define

$$\Sigma_{1|2}(0) = \Sigma_{11}(0) - \Sigma_{12}(0)\Sigma_{22}^{-1}(0)\Sigma_{21}(0), \quad M_{21}(0) = \Sigma_{22}^{-1}(0)\Sigma_{21}(0). \quad (\text{A.46})$$

Thus,

$$u_{1,t}|u_{2,t} \sim N(u_{2,t}'M_{21}(0), \Sigma_{1|2}(0)).$$

Now condition on $y_{1,t} = -\mu_1$ which we can express as $u_{1,t}(0) \leq -\mu_1 - x_t' \Phi_{\cdot 1}(0)$:

$$\begin{aligned} & p(u_{1,t}|y_{1,t} = -\mu_1, u_{2,t}(0), \cdot) \\ &= \frac{p_N(u_{1,t}; u_{2,t}'M_{21}(0), \Sigma_{1|2}(0)) \mathbb{I}\{u_{1,t}(0) \leq -\mu_1 - x_t' \Phi_{\cdot 1}(0)\}}{F_N\left(-\frac{\mu_1 + x_t' \Phi_{\cdot 1}(0) + u_{2,t}'M_{21}(0)}{\sqrt{\Sigma_{12}(0)}}\right)}, \end{aligned} \quad (\text{A.47})$$

which is the density of a truncated normal distribution. To generate a draw i conditional on $(\mu, \Phi(0), \Sigma(0))$ from the posterior of the shadow rate, we can proceed as follows:

- (i) Compute $u'_{2,t}(0) = y'_{2,t} - x'_t \Phi_{.2}(0)$.
- (ii) Compute $\Sigma_{1|2}(0)$ and $M_{21}(0)$ using (A.46).
- (iii) Draw $u^i_{1,t} \sim p(u_{1,t} | y_{1,t} = -\mu_1, u_{2,t}(0), \cdot)$.
- (iv) Compute $y^*_{1,t} = x'_t \Phi_{.1}(0) + u^i_{1,t}$.
- (v) Let $\tilde{y}^*_{1,t} = \mu_1 + y^*_{1,t}$.

Because x_t depends on $y_{1,t-1:t-p}$ instead of $y^*_{1,t-1:t-p}$ inference on the shadow rate is static and does not require dynamic filtering. The last step shifts the re-centered shadow rate back to its original level.

Stochastic Volatility. In every period, we compute the shadow rate based on the filtered volatilities $\mathbb{E}[D_{ii,t} | Y_{1:t}, \theta]$, adjusting the Σ terms in the homoskedastic expressions to reflect the time variation of the shock volatility: $\Sigma_t(s) = \Phi_\epsilon(s) D_t \Phi_\epsilon(s)'$.

C Sequential Monte Carlo Algorithm

Algorithm 2 (Generic SMC Algorithm)

1. **Initialization.** ($n = 0$ and $\phi_0 = 0$.) Draw the initial particles from the prior: $\theta_1^i \stackrel{iid}{\sim} p(\theta)$ and $W_1^i = 1$, $i = 1, \dots, N$.
2. **Recursion.** For $n = 1, \dots, N_\phi$,

- (a) **Correction.** Reweight the particles from stage $n - 1$ by defining the incremental weights

$$\tilde{w}_n^i = \frac{p_n(Y | \theta_{n-1}^i)}{p_{n-1}(Y | \theta_{n-1}^i)} \quad (\text{A.48})$$

and the normalized weights

$$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N} \sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}, \quad i = 1, \dots, N. \quad (\text{A.49})$$

- (b) **Selection (Optional).** Resample the swarm of particles, $\{\theta_{n-1}^i, \tilde{W}_n^i\}_{i=1}^N$, and denote resampled particles by $\{\hat{\theta}_n^i, W_n^i\}_{i=1}^N$, where $W_n^i = 1$ for all i .

(c) **Mutation.** Starting from $\hat{\theta}_n^i$, propagate the particles $\{\hat{\theta}_n^i, W_n^i\}$ via N_{MH} steps of a Metropolis-Hastings (MH) algorithm with transition density $K_n(\theta|\tilde{\theta}; \zeta_n)$ and stationary distribution $\pi_n(\theta)$. Note that the weights are unchanged, and denote the mutated particles by $\{\theta_n^i, W_n^i\}_{i=1}^N$.

3. The importance sampling approximation of the posterior mean $\mathbb{E}_\pi[h(\theta)]$ is given by:

$$\bar{h}_{N_\phi, N} = \sum_{i=1}^N h(\theta_{N_\phi}^i) W_{N_\phi}^i. \quad (\text{A.50})$$

Moreover, the marginal likelihood can be approximate by

$$\hat{p}(Y) = \prod_{n=1}^{N_\phi} \left(\frac{1}{N} \sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i \right). \quad (\text{A.51})$$

We initialize the SMC algorithm for $n = 0$ with draws from the prior density, obtained from a rejection sampler that accounts for the indicator functions $f_u(\theta)$ and $f_s(\theta)$ in the definition of $p(\theta, \lambda)$ in (32).

The correction step is a classic importance sampling step, in which the particle weights are updated to reflect the stage n distribution $\pi_n(\theta)$. The selection step is executed if the variance of the particle weights, transformed into an effective particle sample size

$$\widehat{ESS}_n = N / \left(\frac{1}{N} \sum_{i=1}^N (\tilde{W}_n^i)^2 \right), \quad (\text{A.52})$$

falls below the threshold $N/2$. This threshold rule balances the trade-off between adding noise to the Monte Carlo approximation through resampling and inaccuracies due to very uneven particle weights. The resampling is implemented through systematic resampling; see the textbooks by Liu (2001) or Cappé, Moulines, and Ryden (2005).

The mutation step changes the particle values. In the absence of the mutation step, the particle values would be restricted to the set of values drawn in the initial stage from the prior distribution. This would clearly be inefficient, because the prior distribution is typically a poor proposal distribution for the posterior in an importance sampling algorithm. As the algorithm cycles through the N_ϕ stages, the particle values successively adapt to the shape of the posterior distribution. This is the key difference between SMC and classic importance sampling. The transition kernel $K_n(\theta|\tilde{\theta}; \zeta_n)$ is designed to have the following invariance

property:

$$\pi_n(\theta_n) = \int K_n(\theta_n|\hat{\theta}_n; \zeta_n)\pi_n(\hat{\theta}_n)d\hat{\theta}_n. \quad (\text{A.53})$$

Thus, if $\hat{\theta}_n^i$ is a draw from π_n , then so is θ_n^i . The mutation step is implemented by using multiple steps of a Metropolis-Hastings (MH) algorithm. The vector ζ_n summarizes the tuning parameters of the MH algorithm.¹⁷

D Data Sources

All series are obtained from the economic research database of the Federal Reserve Bank of St. Louis (FRED). Real GDP is *gdpc1*, real potential GDP is *gdppot*, the personal consumption expenditure deflator is *dpcerd3q086sbea*, and the federal funds rate is *fedfunds*. We average the monthly rates to time-aggregate the interest rates to quarterly frequency.

E Additional Empirical Results

Table A-1 provides summary statistics for the posterior distribution of the model parameters.

In Figures A-2 and A-3 we compare prior (grey) and posterior (green) uncertainty about the impulse responses generated conditional on the states of the economy in 1999:Q1 and 2009:Q1, respectively. Our prior distribution allows for a wide variety of responses to the three shocks. While the prior distribution of the impact effect is constrained by the sign restrictions imposed on α_S^Δ , β_D , and γ_D , the wide bands for $h \geq 2$ indicate that the prior leaves the sign of the dynamic effect largely unconstrained. The posterior distribution of the impulse responses is substantially more concentrated, reflecting the sample information about the effects of the three shocks.

Figure A-4 plots the estimated structural shocks conditional on the posterior mean. The structural shocks are generated by converting the reduced-form residuals \hat{u}_t using the matrix $\Phi_\epsilon(s)$. During the ELB period there is uncertainty about the realization of the reduced-form residuals, which is handled in the same way as uncertainty about the shadow rate; see Section B.3.

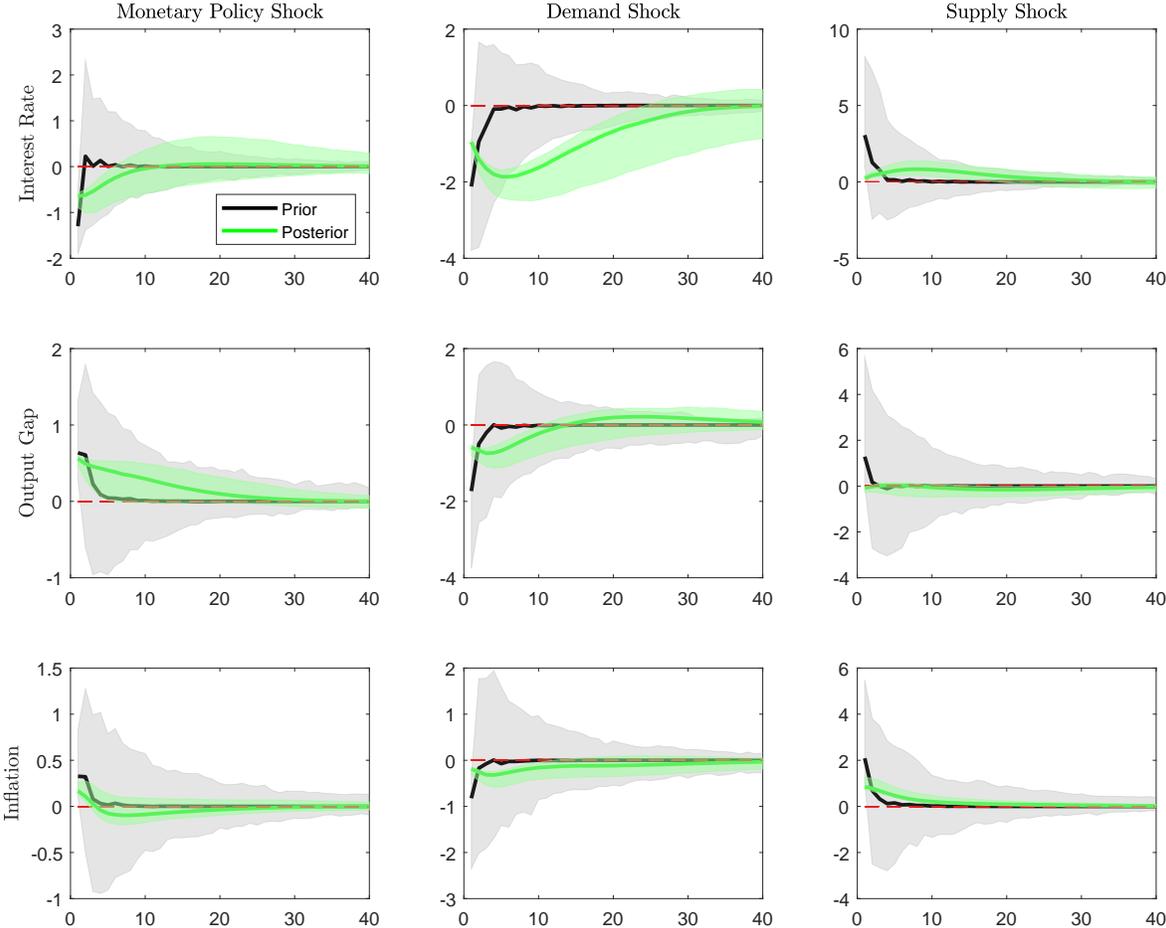
¹⁷The tuning of the mutation step follows Herbst and Schorfheide (2015).

Table A-1: Posterior Distribution

Parameter	Mean	HPD Low	HPD High	Parameter	Mean	HPD Low	HPD High
μ_1	5.32			$\Phi_{10,2}$	0	-0.05	0.03
μ_2	-0.94			$\Phi_{11,2}$	-0.02	-0.04	0.01
μ_3	2.56			$\Phi_{12,2}$	-0.01	-0.05	0.02
ψ_π	0.45	0.24	0.72	$\Phi_{1,3}$	0.08	-0.02	0.15
ψ_z	1.43	1.16	1.63	$\Phi_{2,3}$	0.02	-0.04	0.14
α_S^Δ	2.99	2.12	3.63	$\Phi_{3,3}$	0.92	0.84	1.03
β_D	0.12	0.02	0.18	$\Phi_{4,3}$	0	-0.11	0.12
$-\gamma_D$	0.78	0.56	0.94	$\Phi_{5,3}$	-0.02	-0.12	0.06
$B_{1,1}$	1.08	0.91	1.21	$\Phi_{6,3}$	-0.08	-0.18	-0.01
$B_{2,1}$	-1.15	-1.39	-0.81	$\Phi_{7,3}$	-0.03	-0.08	0.05
$B_{3,1}$	-0.3	-0.56	0.04	$\Phi_{8,3}$	-0.05	-0.09	0.01
$B_{4,1}$	-0.18	-0.31	-0.03	$\Phi_{9,3}$	-0.02	-0.07	0.06
$B_{5,1}$	-0.05	-0.23	0.05	$\Phi_{10,3}$	-0.01	-0.05	0.02
$B_{6,1}$	-0.01	-0.11	0.14	$\Phi_{11,3}$	0	-0.04	0.03
$B_{7,1}$	0.02	-0.09	0.1	$\Phi_{12,3}$	-0.01	-0.05	0.02
$B_{8,1}$	0.02	-0.05	0.08	$\Phi_{12,1}^\Delta$	-0.05	-0.16	0.12
$B_{9,1}$	0.11	-0.01	0.18	$\Phi_{12,2}^\Delta$	-0.36	-0.52	-0.21
$B_{10,1}$	0.02	-0.04	0.09	D_{11}	0.6	0.43	0.71
$B_{11,1}$	0.01	-0.03	0.06	D_{22}	0.44	0.35	0.53
$B_{12,1}$	0.02	-0.03	0.08	D_{33}	1.64	0.68	2.2
$\Phi_{1,2}$	0.1	0.02	0.22	λ_ϕ	0.81	0.74	0.92
$\Phi_{2,2}$	0.98	0.87	1.07	λ_0	11.87	5.88	18.41
$\Phi_{3,2}$	0.03	-0.12	0.15	λ_l	3.25	2.46	4.07
$\Phi_{4,2}$	-0.06	-0.18	0.02	$\rho_{\zeta,1}$	0.89	0.83	0.97
$\Phi_{5,2}$	-0.03	-0.13	0.07	$\rho_{\zeta,2}$	0.79	0.67	0.89
$\Phi_{6,2}$	-0.02	-0.09	0.09	$\rho_{\zeta,3}$	0.88	0.84	0.95
$\Phi_{7,2}$	-0.05	-0.1	0.02	ς_1	0.29	0.18	0.42
$\Phi_{8,2}$	-0.02	-0.06	0.02	ς_2	0.3	0.11	0.45
$\Phi_{9,2}$	-0.04	-0.09	0.03	ς_3	0.29	0.18	0.36

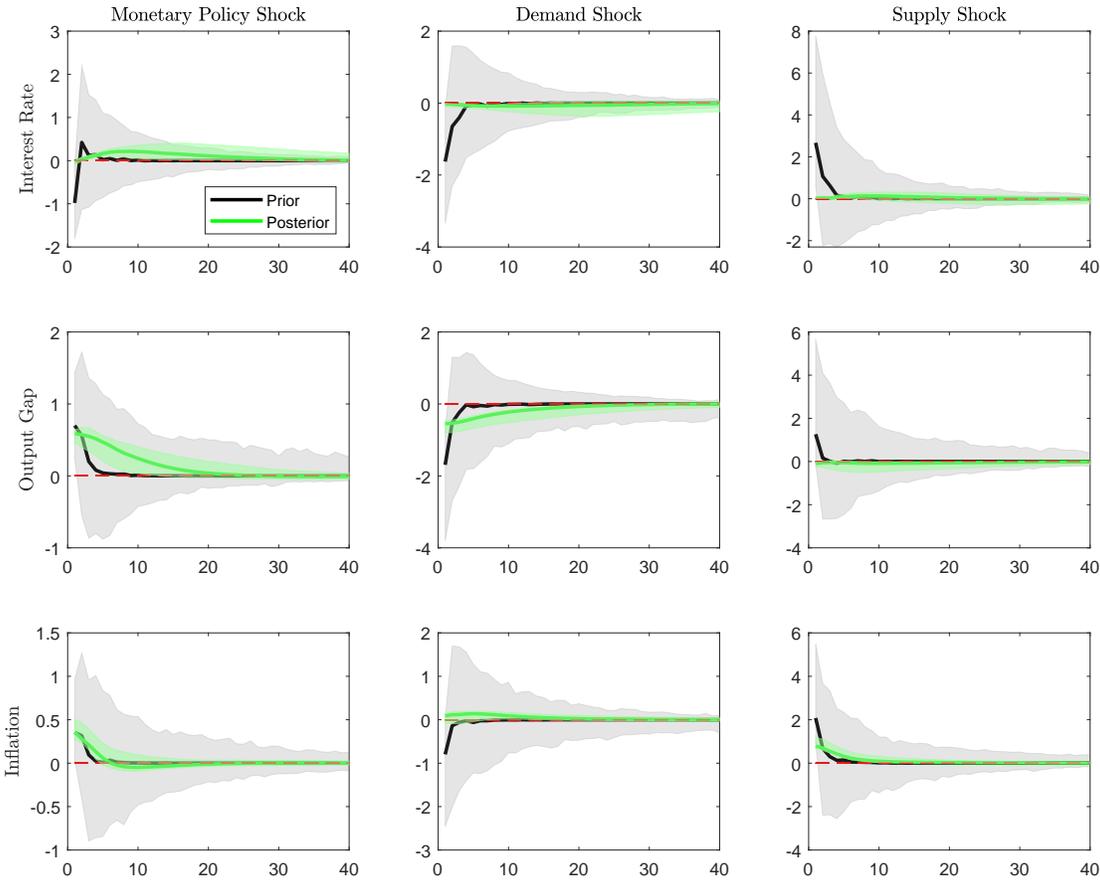
Notes: HPD Low and High are the 5th and 95th percentiles of the (marginal) posterior distribution. The following hyperparameters are fixed: $\lambda_d = 4$, $\lambda_\Delta = 16$, and $\lambda_l = \infty$, meaning we are fixing the level parameters μ_1 , μ_2 , and μ_3 .

Figure A-2: Responses to Minus-Two-Standard-Deviation Shocks with Flexible Regimes and Prior/Posterior Parameter Uncertainty in 1999:Q1



Notes: Prior median (black) and 90% credible interval (grey); posterior median (green) and 90% credible interval (light green).

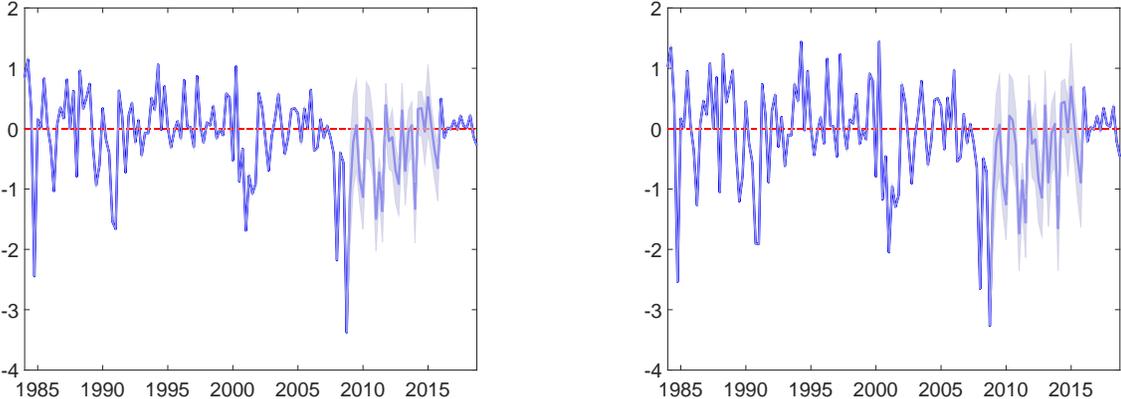
Figure A-3: Responses to Minus-Two-Standard-Deviation Shocks with Flexible Regimes and Prior/Posterior Parameter Uncertainty in 2009:Q1



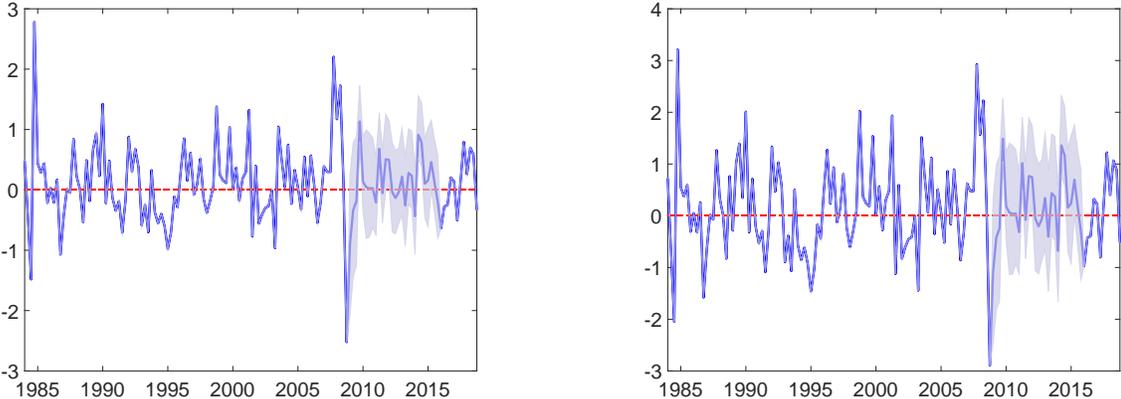
Notes: Prior median (black) and 90% credible interval (grey); posterior median (green) and 90% credible interval (light green).

Figure A-4: Historical Shocks

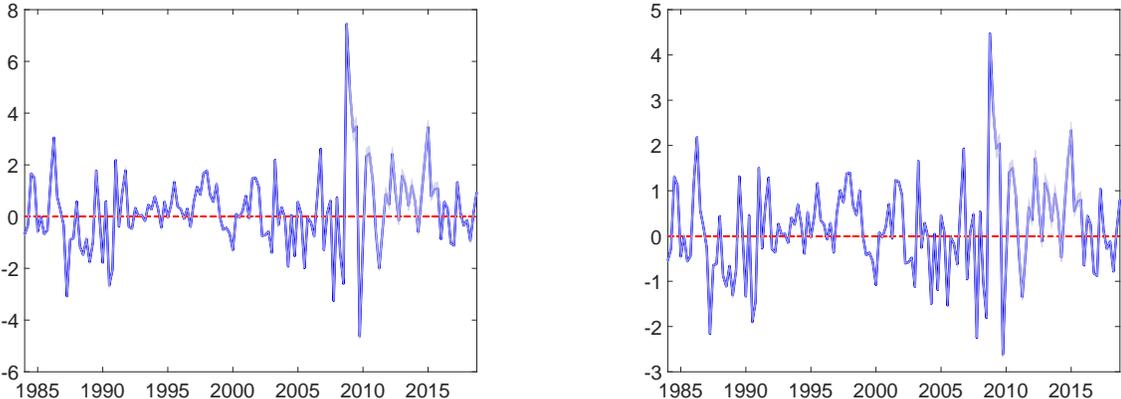
Monetary Policy Shock



Demand Shock



Supply Shock



Level

Standardized

Notes: Shocks are computed conditional on posterior mean $\bar{\theta}$. During the ELB period the grey bands indicate 90% equal-tail-probability bands that reflect uncertainty about the realization of the shocks.