

# Identification at the Zero Lower Bound

Sophocles Mavroeidis\*

University of Oxford and INET

October 15, 2020

## Abstract

I show that the Zero Lower Bound (ZLB) on interest rates can be used to identify the causal effects of monetary policy. Identification depends on the extent to which the ZLB limits the efficacy of monetary policy. I develop a general econometric methodology for the identification and estimation of structural vector autoregressions (SVARs) with an occasionally binding constraint. The method provides a simple way to test the efficacy of unconventional policies, modelled via a ‘shadow rate’. I apply this method to U.S. monetary policy using a three-equation SVAR model of inflation, unemployment and the federal funds rate. I reject the null hypothesis that unconventional monetary policy has no effect at the ZLB, but find some evidence that it is not as effective as conventional monetary policy.

Keywords: SVAR, censoring, coherency, partial identification, monetary policy, shadow rate

JEL codes: C1, C3, C5, E4, E5

## 1 Introduction

The zero lower bound (ZLB) on nominal interest rates has arguably been a challenge for policy makers and researchers of monetary policy. Policy makers have had to resort to so-called unconventional policies, such as quantitative easing or forward guidance, which had previously been largely untested. Researchers have to use new theoretical and empirical methodologies to analyze macroeconomic models when the ZLB binds. So, the ZLB is generally viewed as a problem or at least a nuisance. This paper proposes to turn this problem on its head to solve another long-standing question in macroeconomics: the identification of the causal effects of monetary policy on the economy.

---

\*sophocles.mavroeidis@economics.ox.ac.uk. This research is funded by the European Research Council via Consolidator grant number 647152. I would like to thank Guido Ascari, Andrea Ferrero, Jim Hamilton, Frank Schorheide, Francesco Zanetti and seminar participants at the Chicago Fed, the Bank of Japan, the NBER Summer Institute, UCL, the DNB, University of Pennsylvania, Pennsylvania State University and the University of Oxford for useful comments and discussion, as well as Mishel Ghassibe, Lukas Freund, Shangshang Li, and Patrick Vu for research assistance.

The intuition is as follows. If the ZLB limits the ability of policy makers to react to macroeconomic shocks,<sup>1</sup> the response of the economy to shocks will change when the policy instrument hits the ZLB. Because the difference in the behavior of macroeconomic variables across the ZLB and non-ZLB regimes is only due to the impact of monetary policy, the switch across regimes provides information about the causal effects of policy. In the extreme case that monetary policy completely shuts down during the ZLB regime, either because policy makers do not use alternative (unconventional) policy instruments, or because such instruments turn out to be completely ineffective, the only difference in the behavior of the economy across regimes is due to the impact of (conventional) policy during the unconstrained regime. Therefore, the ZLB identifies the causal effect of policy during the unconstrained regime. If monetary policy remains partially effective during the ZLB regime, e.g., through the use of unconventional policy instruments, then the difference in the behavior of the economy across regimes will depend on the difference in the effectiveness of conventional and unconventional policies. In this case, we obtain only partial identification of the causal effects of monetary policy, but we can still get informative bounds on the relative efficacy of unconventional policy. In the other extreme case that unconventional policy is as effective as conventional policy, there is no difference in the behavior of the economy across regimes, and we have no additional information to identify the causal effects of policy. However, we can still test this so-called ZLB irrelevance hypothesis (Debortoli et al., 2019) by testing whether the reaction of the economy to shocks is the same across the two regimes.

There are similarities between identification via occasionally binding constraints and identification through heteroskedasticity (Rigobon, 2003), or more generally, identification via structural change (Magnusson and Mavroeidis, 2014). That literature showed that the switch between different regimes generates variation in the data that identifies parameters that are constant across regimes. For example, an exogenous shift in a policy reaction function or in the volatility of shocks identifies the transmission mechanism, provided the latter is unaffected by the policy shift.<sup>2</sup> An instrumental variables interpretation of identification via structural change is that regime indicators (dummy variables) can be used as instruments for the endogenous policy variable. When the switch from one regime to another is exogenous, regime indicators are valid instruments, and the methodology in Magnusson and Mavroeidis (2014) is applicable.<sup>3</sup> However, regimes induced by occasionally binding constraints are not exogenous – whether the ZLB binds or not clearly depends on the structural shocks, so regime indicators cannot be used as instruments in the usual way, and a new methodology is needed to analyze these models.

In this paper, I show how to control for the endogeneity in regime selection and obtain identification

---

<sup>1</sup>As argued, for example, by Eggertsson and Woodford (2003) who concluded that the ZLB “does represent an important constraint on what monetary policy can achieve.”

<sup>2</sup>Examples of such policy regime shifts are a switch from passive to active US monetary policy (Clarida et al. (2000)), or changes in the inflation target (Cogley and Sbordone, 2008).

<sup>3</sup>Magnusson and Mavroeidis (2014) deal with the additional complication that the timing and number of regimes might be unknown.

in structural vector autoregressions (SVARs).<sup>4</sup> The methodology is parametric and likelihood-based, and the analysis is similar to the well-known Tobit model (Tobin, 1958). More specifically, the methodological framework builds on the early microeconometrics literature on simultaneous equations models with censored dependent variables, see Amemiya (1974), Lee (1976), Blundell and Smith (1994), and the more recent literature on dynamic Tobit models, see Lee (1999), and particle filtering, see Pitt and Shephard (1999).

A further contribution of this paper is a general methodology to estimate reduced-form VARs with a variable subject to an occasionally binding constraint. This is a necessary starting point for SVAR analysis that uses any of the existing popular identification schemes, such as short- or long-run restrictions, sign restrictions, or external instruments. In the absence of any constraints, reduced-form VARs can be estimated consistently by Ordinary Least Squares (OLS), which is Gaussian Maximum Likelihood, or its corresponding Bayesian counterpart, and inference is fairly well-established. However, it is well-known that OLS estimation is inconsistent when the data is subject to censoring or truncation, see, e.g., Greene (1993) for a textbook treatment. So, it is not possible to estimate a VAR consistently by OLS using any sample that includes the ZLB, or even using (truncated) subsamples when the ZLB is not binding (because of selection bias), as was pointed out by Hayashi and Koeda (2019). It is not possible to impose the ZLB constraint using Markov switching models with exogenous regimes, as in Liu et al. (2019), because exogenous Markov-switching cannot guarantee that the constraint will be respected with probability one, and also does not account for the fact that the switch from one regime to the other depends on the structural shocks.<sup>5</sup> Finally, it is also not possible to perform consistent estimation and valid inference on the VAR (i.e., error bands with correct coverage on impulse responses), using externally obtained measures of the shadow rate, such as the one proposed by Wu and Xia (2016), as the estimation error in any such measures is not asymptotically negligible and is typically highly autocorrelated.<sup>6</sup> In other words, shadow rate estimates are subject to large and persistent measurement error that is not accounted for if they are treated as known in subsequent analysis. The methodology developed in this paper allows for the presence of a shadow rate, estimates of which can be obtained, but more importantly, it fully accounts for the impact of sampling uncertainty in the estimation of the shadow rate on inference about the structural parameters such as impulse responses. Therefore, the paper fills an important gap in the literature, as it provides the requisite methodology

---

<sup>4</sup>There is a related literature on Dynamic Stochastic General Equilibrium (DSGE) models with a ZLB, see, e.g., Fernández-Villaverde et al. (2015), Guerrieri and Iacoviello (2015), Aruoba et al. (2017), Kulish et al. (2017) and Aruoba et al. (2020). The papers in this literature do not point out the implications of the ZLB for identification of monetary policy shocks.

<sup>5</sup>Hayashi and Koeda (2019) develop a VAR model with endogenous regime switching in which the policy variables that are subject to a lower bound, the short-term interest rate and excess reserves, are modelled using Tobit regressions. A key difference of their methodology from the one developed here is that they impose recursive identification of monetary policy shocks, which the present paper shows to be an overidentifying, and hence testable, restriction. Moreover, their model does not include shadow rates.

<sup>6</sup>See also Rossi (2019) for a comprehensive discussion of the challenges posed by the ZLB for the estimation of structural VARs.

to implement any of the existing identification schemes.<sup>7</sup>

Identification of the causal effects of policy by the ZLB does not require that the policy reaction function be stable across regimes. However, inference on the efficacy of unconventional policy, or equivalently, the causal effects of shocks to the shadow rate over the ZLB period, obviously depends on whether or not the reaction function remains the same across regimes. For example, an attenuation of the causal effects of policy over the ZLB period may indicate that unconventional policy is only partially effective, but it is also consistent with unconventional policy being less active (during ZLB regimes) than conventional policy (during non-ZLB regimes). This is a fundamental identification problem that is difficult to overcome without additional information, such as measures of unconventional policy stance, or additional identifying assumptions, such as parametric restrictions or external instruments. This can be done using the methodology developed in this paper.

The structure of the paper is as follows. Section 2 presents the main identification results of the paper in the context of a static bivariate simultaneous equations model with a limited dependent variable subject to a lower bound. Section 3 generalizes the analysis to a SVAR with an occasionally binding constraint and discusses identification, estimation and inference. Section 4 provides an application to a three-equation SVAR in inflation, unemployment and the Federal funds rate from Stock and Watson (2001). Using a sample of post-1960 quarterly US data, I find some evidence that the ZLB is empirically relevant, and that unconventional policy is only partially effective. Proofs, computational details and simulations are given in the Appendix at the end. Additional supporting material is provided in a Supplementary Appendix available online.

## 2 Simultaneous equations model

I first illustrate the idea using a simple bivariate simultaneous equations model (SEM), which is both analytically tractable and provides a direct link to the related microeconometrics literature. To make the connection to the leading application, I will motivate this using a very stylized economy without dynamics in which the only outcome variable is inflation  $\pi_t$  and the (conventional) policy instrument is the short-term nominal interest rate,  $r_t$ . In addition to the traditional interest rate channel, the model allows for an ‘unconventional monetary policy’ channel that can be used when the conventional policy instrument hits the ZLB. An example of such a policy is quantitative easing (QE), in the form

---

<sup>7</sup>A very recent paper by Aruoba et al. (2020) also studies SVARs with occasionally binding constraints, but does not focus on the implications of those constraints for identification. Another difference is that the private sector part of the SVAR is modelled in  $\Phi$ -form ( $Y_t = \Phi\varepsilon_t + lags$ ), namely, in terms of the decision rules of the agents, while I use the  $A$ -form ( $AY_t = \varepsilon_t + lags$ ), so that the decision rules are obtained after solving the model. Both models have regime-dependent piecewise linear and continuous decision rules, but they are not equivalent, because, unlike linear SVARs, the  $\Phi$  and  $A$  forms are not equivalent when there are occasionally binding constraints. Finally, their SVAR does not include shadow rates, so it cannot nest commonly used forward guidance policy rules, such as Reifschneider and Williams (2000), or Debortoli et al. (2019), or the case of ZLB irrelevance.

of long-term asset purchases by the central bank. Here I discuss a simple model of QE.<sup>8</sup>

Abstracting from dynamics and other variables, the equation that links inflation to monetary policy is given by

$$\pi_t = c + \beta \left( r_t - r^n + \varphi \hat{b}_{L,t} \right) + \varepsilon_{1t}, \quad (1)$$

where  $c$  is a constant,  $r^n$  is the neutral rate,  $\hat{b}_{L,t}$  is the amount of long-term bonds held by the private sector in log-deviation from its steady state, and  $\varepsilon_{1t}$  is an exogenous structural shock unrelated to monetary policy. Equation (1) can be obtained from a model of bond-market segmentation, as in Chen et al. (2012), where a fraction of households is constrained to invest only in long-term bonds, see the Supplementary Appendix for more details. In such a model, the parameter  $\varphi$  that determines the effectiveness of QE is proportional to the fraction of constrained households and the elasticity of the term premium with respect to asset holdings, both of which are assumed to be outside the control of the central bank.

The nominal interest rate is set by a Taylor rule subject to the ZLB constraint, namely,

$$r_t = \max(r_t^*, 0), \quad (2a)$$

$$r_t^* = r^n + \gamma \pi_t + \varepsilon_{2t}, \quad (2b)$$

where  $r_t^*$  represents the desired target policy rate, and  $\varepsilon_{2t}$  is a monetary policy shock. When  $r_t^*$  is negative, it is unobserved. The unobserved  $r_t^*$  will be referred to as the ‘shadow rate’, and it represents the desired policy stance prescribed by the Taylor rule in the absence of a binding ZLB constraint.

Suppose that QE is activated only when the conventional policy instrument  $r_t$  hits the ZLB,<sup>9</sup> and follows the same policy rule (2b), up to a factor of proportionality  $\alpha$ , i.e.,

$$\hat{b}_{L,t} = \min(\alpha r_t^*, 0).$$

Substituting for  $\hat{b}_{L,t}$  in eq. (1), we obtain

$$\pi_t = c + \beta (r_t - r^n) + \beta^* \min(r_t^*, 0) + \varepsilon_{1t}, \quad (3)$$

where  $\beta^* = \alpha \varphi \beta$ .

A special case arises when QE is ineffective ( $\varphi = 0$ ), or the monetary authority does not pursue a

---

<sup>8</sup>The more general SVAR model of the next section can also incorporate forward guidance in the form of Reifschneider and Williams (2000) and Debortoli et al. (2019), as shown in Ikeda et al. (2020).

<sup>9</sup>The assumption that QE is only active during the ZLB regime is only made for simplicity, as it is inconsequential for the resulting functional form of the transmission equation. We can let QE be active all the time, and even allow for a different rule for QE above and below the ZLB, i.e.,  $\hat{b}_{L,t} = \alpha \min(r_t^*, 0) + \alpha_1 \max(r_t^*, 0)$ , which can also be written as  $\hat{b}_{L,t} = \alpha r_t^* + (\alpha_1 - \alpha) r_t$ . Then, substituting back into (1) yields an equation that is isomorphic to (3), i.e.,  $\pi_t = c_1 + \bar{\beta} r_t + \bar{\beta}^* \min(r_t^*, 0) + \varepsilon_{1t}$ , with  $\bar{\beta} = \beta(1 + \varphi \alpha_1)$  and  $\bar{\beta}^* = \beta \varphi \alpha$ .

QE policy ( $\alpha = 0$ ), so that eq. (3) becomes

$$\pi_t = c + \beta (r_t - r^n) + \varepsilon_{1t}, \quad (4)$$

and monetary policy is completely inactive at the ZLB.

Another special case of the model given by equations (2) and (3) arises when  $\beta^* = \beta$  in eq. (3). This happens when  $\varphi \neq 0$  and  $\alpha$  is chosen by the monetary authority to be equal to  $1/\varphi$ . This can be done when policy makers know the transmission mechanism in eq. (1) and have no restrictions in setting the policy parameter  $\alpha$  so as to fully remove the impact of the ZLB on conventional policy. In that case, the equation for the outcome variable becomes

$$\pi_t = c + \beta (r_t^* - r^n) + \varepsilon_{1t}. \quad (5)$$

The model given by equations (2) and (5) is one in which monetary policy is completely unconstrained and there is no difference in outcomes across policy regimes. Such models have been put forward by Swanson and Williams (2014), Debortoli et al. (2019) and Wu and Zhang (2019).

The nesting model given by eq. (3) allows the effects of conventional and unconventional policy to differ. This could reflect informational as well as political or institutional constraints that prevent policy makers from calibrating their unconventional policy response to match exactly the policy prescribed by the Taylor rule. For instance, it may be that policy makers do not know the effectiveness of the QE channel  $\varphi$ , or that the scale of asset purchases needed to achieve the desired policy stance during a ZLB regime is too large to be politically acceptable. Such a consideration may be particularly pertinent, for example, in the Eurozone. Importantly, one does not need to take a theoretical stand on this issue, because the methodology that I develop in the paper can accommodate a wide range of possibilities, and, as I demonstrate below, the issue can be studied empirically.

Letting  $\lambda = \alpha\varphi$  and using the fact that  $\min(r_t^*, 0) = r_t^* - r_t$ , eq. (3) can be written equivalently as

$$\pi_t = c + \beta ((1 - \lambda) r_t + \lambda r_t^* - r^n) + \varepsilon_{1t}. \quad (6)$$

If  $\lambda = 1$ , policy is completely unconstrained by the ZLB, while if  $\lambda = 0$ , unconventional policy is completely inactive or ineffective.

To complete the specification of the model, I assume that the structural shocks  $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$  are independently and identically distributed (*i.i.d.*) Normal with covariance matrix  $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2)$ .

**Connection to the microeconometrics literature** Equations (2) and (3) or (6) form a SEM with a limited dependent variable. The special case with  $\beta^* = 0$  in (3), or  $\lambda = 0$  in (6), can be referred to as a *kinked* SEM (KSEM), while the opposite case of  $\beta^* = \beta$  in (3), or  $\lambda = 1$  in (6),

can be called a *censored* SEM (CSEM). Variants of the KSEM model have been studied in the early microeconometrics literature on limited dependent variable SEMs. Amemiya (1974) and Lee (1976) studied multivariate extensions of the well-known Tobit model (Tobin, 1958). Nelson and Olson (1978) argued that the KSEM was less suitable for microeconomic applications than the CSEM, and the latter subsequently became the main focus of the literature (Smith and Blundell, 1986; Blundell and Smith, 1989). Blundell and Smith (1994) studied the unrestricted model using external instruments, so they did not consider the implications of the kink for identification.

One important lesson from the microeconometrics literature is that establishing existence and uniqueness of equilibria in this class of models is non-trivial. Gouriéroux et al. (1980) define a model to be ‘coherent’ if it has a unique solution for the endogenous variables in terms of the exogenous variables, i.e., if there exists a unique reduced form. More recently, the literature has distinguished between existence and uniqueness of solutions using the terms coherency and completeness of the model, respectively (Lewbel, 2007). Establishing coherency and completeness is a necessary first step before we can study identification and estimation.<sup>10</sup>

## 2.1 Identification

Substituting for  $r_t^*$  in (6) using (2), and rearranging, we obtain

$$\pi_t = \tilde{c} + \tilde{\beta}(r_t - r^n) + \tilde{\varepsilon}_{1t}, \text{ where} \quad (7)$$

$$\tilde{\beta} = \frac{(1-\lambda)\beta}{1-\lambda\beta\gamma} = \frac{\beta-\beta^*}{1-\gamma\beta^*}, \quad \tilde{c} = \frac{c}{1-\lambda\beta\gamma}, \text{ and } \tilde{\varepsilon}_{1t} = \frac{\varepsilon_{1t} + \lambda\beta\varepsilon_{2t}}{1-\lambda\beta\gamma}. \quad (8)$$

The system of equations (2) and (7) is now a KSEM, for which the necessary and sufficient condition for coherency and completeness (existence of a unique solution) is  $\tilde{\beta}\gamma < 1$  (Nelson and Olson, 1978). Using (8), the coherency and completeness condition can be expressed in terms of the structural parameters as  $(1-\gamma\beta)/(1-\lambda\gamma\beta) > 0$ , or, equivalently, as  $(1-\gamma\beta)/(1-\gamma\beta^*) > 0$ . This condition evidently restricts the admissible range of the structural parameters.

Under the condition  $\tilde{\beta}\gamma < 1$ , the unique solution of the model can be written as

$$\pi_t = \mu_1 + u_{1t} - \tilde{\beta}D_t(\mu_2 + u_{2t}), \text{ and} \quad (9)$$

$$r_t = \max(\mu_2 + u_{2t}, 0), \quad (10)$$

where  $D_t := 1_{\{r_t=0\}}$  is an indicator (dummy) variable that takes the value 1 when  $r_t$  is on the boundary

---

<sup>10</sup>Despite its importance, the coherency condition (which is the same as existence of an equilibrium) is rarely addressed in the literature on the ZLB because it is generally analytically intractable, see Ascari and Mavroeidis (2020) and the references therein for further discussion of this issue.

and zero otherwise, and

$$\begin{aligned} u_{1t} &= \frac{\tilde{\varepsilon}_{1t} + \tilde{\beta}\varepsilon_{2t}}{1 - \gamma\tilde{\beta}} = \frac{\varepsilon_{1t} + \beta\varepsilon_{2t}}{1 - \gamma\beta}, & u_{2t} &= \frac{\gamma\varepsilon_{1t} + \varepsilon_{2t}}{1 - \gamma\beta}, \\ \mu_1 &= \frac{c}{1 - \gamma\beta}, & \mu_2 &= \frac{\gamma c}{1 - \gamma\beta} + r^n. \end{aligned} \quad (11)$$

Equations (9) and (10) express the endogenous variables  $\pi_t, r_t$  in terms of the exogenous variables  $\varepsilon_{1t}, \varepsilon_{2t}$ , and correspond to the decision rules of the agents in the model. It is clear that those decision rules differ in a world in which the ZLB occasionally binds, which is characterized by  $\tilde{\beta} \neq 0$ , compared to a world in which it never does (i.e., the KSEM), where  $\tilde{\beta} = 0$ . What is important for identification, however, is that in a world in which the ZLB occasionally binds, agents' reaction to shocks differs across regimes, and the difference depends on the parameter  $\tilde{\beta}$ , which from eq. (8), depends on the difference between the impact of conventional and unconventional policies,  $\beta$  and  $\beta^*$ , respectively. I will show that this change provides information that identifies the structural parameters: we get point identification when  $\beta^* = 0$  (the KSEM case), and partial identification when  $\beta^* \neq \beta$ . The identification argument leverages the coefficient on the kink,  $\tilde{\beta}$ , in the 'incidentally kinked' regression (9), which is identified by a variant of the well-known Heckit method (Heckman, 1979). I will sketch out the argument below, and provide more details for the full SVAR model in the next section.

### 2.1.1 Identification of the KSEM

Recall that in the KSEM model  $\tilde{\beta} = \beta$ . Consider the estimation of  $\beta$  in (4) from a regression of  $\pi_t$  on  $r_t$  using only observations above the ZLB,

$$E(\pi_t | r_t, r_t > 0) = c + \beta(r_t - r^n) + \rho \left( r_t - \mu_2 + \tau \frac{\phi(a)}{1 - \Phi(a)} \right), \quad a = \frac{-\mu_2}{\tau}, \quad (12)$$

where  $\rho = \text{cov}(u_{1t}, u_{2t}) / \tau^2 - \beta = \gamma\sigma_1^2(1 - \gamma\beta) / (\gamma^2\sigma_1^2 + \sigma_2^2)$ ,  $\tau = \sqrt{\text{var}(u_{2t})}$ , and  $\phi(\cdot)$ ,  $\Phi(\cdot)$  are standard Normal density and distribution functions, respectively. The coefficient  $\rho$  is the bias in the estimation of  $\beta$  from the truncated regression (12). Now, the mean of  $\pi_t$  using the observations at the ZLB is

$$E(\pi_t | r_t = 0) = c - \beta r^n + \rho \tau \frac{\phi(a)}{\Phi(a)}. \quad (13)$$

Next, observe that  $\mu_2$ ,  $\tau$  and hence  $\phi(a)/\Phi(a)$  can be estimated from the Tobit regression (10). Therefore, we can recover the bias  $\rho$  and identify  $\beta$ . A simple way to implement this is the control function approach (Heckman, 1978). Let

$$h_t(\mu_2, \tau) := (1 - D_t)(r_t - \mu_2) - D_t \frac{\tau \phi(a)}{\Phi(a)},$$



and run the regression

$$E(\pi_t | r_t) = c_1 + \beta r_t + \rho h_t(\mu_2, \tau), \quad (14)$$

where  $c_1 = c - \beta r^n$  is an unrestricted intercept. The rank condition for the identification of  $\beta$  is simply that the regressors in (14) are not perfectly collinear. This holds if and only if  $0 < \Pr(D_t = 1) < 1$ . So, as long as some but not all the observations are at the boundary, the model is generically identified.

### 2.1.2 Partial identification of the unrestricted SEM

The discussion of the previous subsection shows that  $\tilde{\beta}$  is identified from the kink in the reduced-form equation for  $\pi_t$  (9). From (8), it follows that  $\beta, c$  and  $\varepsilon_{1t}$  are underidentified unless  $\lambda$  is known.

Now, consider the restriction  $\lambda \in [0, 1]$ . If we interpret  $\lambda$  as a measure of the efficacy of unconventional policy, this restriction implies that unconventional policy is neither counter- nor over-productive. Specifically, it says that the effect of QE on  $\pi_t$ , which is given by the coefficient  $\beta^* = \lambda\beta$ , cannot be of a different sign or of higher magnitude (in absolute value) than the effect of conventional policy. I will now discuss the implications of this restriction for the identification of the structural parameters.

First, we have already established that  $\beta$  is completely unidentified when  $\lambda = 1$  (which corresponds to the CSEM). From the definition of  $\tilde{\beta}$  in (8) and the fact that  $1 - \lambda\beta\gamma \neq 0$  from the coherency (existence) condition, it follows that  $\lambda = 1$  implies  $\tilde{\beta} = 0$ . So, when  $\tilde{\beta} = 0$ ,  $\beta$  is completely unidentified. It remains to see what happens when  $\tilde{\beta} \neq 0$ . Let  $\omega_{ij} := E(u_{it}u_{jt})$  denote the variances and covariance between the reduced form errors, which are identifiable, and define the coefficient of the regression of  $u_{2t}$  on  $u_{1t}$  by  $\gamma_0 := \omega_{12}/\omega_{11}$ .<sup>11</sup> In the Appendix, I prove the following bounds

$$\begin{aligned} & \text{if } \tilde{\beta} = 0 \text{ or } \tilde{\beta}\gamma_0 < 0, \text{ then } \beta \in \Re; \text{ otherwise} \\ & \text{if } \omega_{12} = \gamma_0 = 0, \text{ then } \beta \in (-\infty, \tilde{\beta}] \text{ if } \tilde{\beta} < 0 \text{ or } \beta \in [\tilde{\beta}, \infty) \text{ if } \tilde{\beta} > 0; \\ & \text{if } 0 < \tilde{\beta}\gamma_0 \leq 1, \text{ then } \beta \in \left[\frac{1}{\gamma_0}, \tilde{\beta}\right] \text{ if } \tilde{\beta} < 0 \text{ or } \beta \in \left[\tilde{\beta}, \frac{1}{\gamma_0}\right] \text{ if } \tilde{\beta} > 0; \\ & \text{if } \tilde{\beta}\gamma_0 > 1, \text{ then } \lambda < 0. \end{aligned} \quad (15)$$

In words, when the coefficient on the kink in the reduced form,  $\tilde{\beta}$ , is different from zero, and it is of the same sign as the coefficient of the regression of  $r_t$  on  $\pi_t$  over the uncensored observations,  $\gamma_0$ , and if  $\tilde{\beta}\gamma_0 \leq 1$ , then we can identify both the sign of the causal effect of  $r_t$  on  $\pi_t$  and get bounds on (the absolute value of) its magnitude. In those cases, the identified coefficient  $\tilde{\beta}$  is an attenuated measure of the true causal effect  $\beta$ . Moreover,  $\tilde{\beta}\gamma_0 > 1$  implies that the coefficient on the shadow rate  $r_t^*$  is of the opposite sign than the coefficient on the observed variable  $r_t$ , i.e., unconventional policy has the opposite effect of the conventional one. That could be interpreted as saying that unconventional policy is counterproductive.

<sup>11</sup>  $\gamma_0$  can also be interpreted as the value the reaction function coefficient  $\gamma$  in (2) would take if  $\beta = 0$ , i.e., the value corresponding to a Choleski identification scheme where  $r_t$  is placed last.

Finally, the hypothesis that the bound is “empirically irrelevant” in the sense that unconventional policy fully removes any constraints on what policy can achieve is equivalent to  $\lambda = 1$ . From the above discussion, it is clear that this hypothesis is testable:  $\lambda = 1$  implies that  $\tilde{\beta} = 0$ , so  $\tilde{\beta} \neq 0$  implies  $\lambda \neq 1$ . Therefore, rejecting  $\tilde{\beta} = 0$  unambiguously implies that the bound is empirically relevant.<sup>12</sup> In fact, the parameter  $\lambda$  is also partially identified, and the identified set can be computed analytically, see Appendix A.1.1.

### 3 SVAR with an occasionally binding constraint

I now develop the methodology for identification and estimation of SVARs with an occasionally binding constraint. Let  $Y_t = (Y'_{1t}, Y_{2t})'$  be a vector of  $k$  endogenous variables, partitioned such that the first  $k - 1$  variables  $Y_{1t}$  are unrestricted and the  $k$ th variable  $Y_{2t}$  is bounded from below by  $b$ .<sup>13</sup> Define the latent process  $Y_{2t}^*$  that is only observed, and equal to  $Y_{2t}$ , whenever  $Y_{2t} > b$ . If  $Y_{2t}$  is a policy instrument,  $Y_{2t}^*$  can be thought of as the ‘shadow’ instrument that measures the desired policy stance. The  $p$ th-order SVAR model is given by the equations

$$A_{11}Y_{1t} + A_{12}Y_{2t} + A_{12}^*Y_{2t}^* = B_{10}X_{0t} + \sum_{j=1}^p B_{1,j}Y_{t-j} + \sum_{j=1}^p B_{1,j}^*Y_{2,t-j}^* + \varepsilon_{1t}, \quad (16)$$

$$A_{22}^*Y_{2t}^* + A_{22}Y_{2t} + A_{21}Y_{1t} = B_{20}X_{0t} + \sum_{j=1}^p B_{2,j}Y_{t-j} + \sum_{j=1}^p B_{2,j}^*Y_{2,t-j}^* + \varepsilon_{2t}, \quad (17)$$

$$Y_{2t} = \max(Y_{2t}^*, b),$$

for  $t \geq 1$  given a set of initial values  $Y_{-s}, Y_{2,-s}^*$ , for  $s = 0, \dots, p - 1$ , and  $X_{0t}$  are exogenous and predetermined variables.

Equation (17) can be interpreted as a policy reaction function because it determines the desired policy stance  $Y_{2t}^*$ . Similarly,  $\varepsilon_{2t}$  is the corresponding policy shock. The above model is a dynamic SEM. Two important differences from a standard SEM are the presence of (i) latent lags amongst the predetermined variables on the right-hand side, which complicates estimation; and (ii) the contemporaneous value of  $Y_{2t}$  in the policy reaction function (17), which allows it to vary across ZLB and non-ZLB regimes. The presence of latent lags  $Y_{2,t-j}^*$  in the policy rule (17) is particularly useful because it allows the model to incorporate forward guidance.<sup>14</sup>

Collecting all the observed predetermined variables  $X_{0t}, Y_{t-1}, \dots, Y_{t-p}$  into a vector  $X_t$ , and the latent lags  $Y_{2,t-1}^*, \dots, Y_{2,t-p}^*$  into  $X_t^*$ , and similarly for their coefficients, the model can be written

<sup>12</sup>The converse is not true since  $\tilde{\beta} = 0$  can also arise from  $\beta = 0$  even when  $\lambda \neq 1$ .

<sup>13</sup>The lower bound does not need to be constant. All we need is to observe the periods in which the economy is at the ZLB regime.

<sup>14</sup>Reifschneider and Williams (2000) and Debortoli et al. (2019) provide examples of policy rules with forward guidance that are nested in (17), see the Supplementary Appendix for details.

compactly as:

$$\begin{pmatrix} A_{11} & A_{12}^* & A_{12} \\ A_{21} & A_{22}^* & A_{22} \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t}^* \\ Y_{2t} \end{pmatrix} = BX_t + B^*X_t^* + \varepsilon_t, \quad (18)$$

$$Y_{2t} = \max \{Y_{2t}^*, b\}.$$

The vector of structural errors  $\varepsilon_t$  is assumed to be *i.i.d.* Normally distributed with zero mean and identity covariance.

In the previous section, we defined the KSEM as a special case of the general model, where  $Y_{2t}^* < b$  has no (contemporaneous) impact on  $Y_{1t}$ . In the dynamic setting, it feels natural to define the corresponding ‘kinked SVAR’ model (KSVAR) as a model in which  $Y_{2t}^*$  has neither contemporaneous nor dynamic effects. Therefore, the KSVAR obtains as a special case of (18) when both  $A_{12}^* = 0$ , and  $B^* = 0$ , which corresponds to a situation in which the bound is fully effective in constraining what policy can achieve at all horizons.

The opposite extreme to the KSVAR is the censored SVAR model (CSVAR). Again, unlike the CSEM, which only characterizes contemporaneous effects, the idea of a CSVAR is to impose the assumption that the constraint is irrelevant at all horizons. So, it corresponds to a fully unrestricted linear SVAR in the latent process  $(Y_{1t}', Y_{2t}^*)'$ . This is a special case of (18) when both  $A_{12} = 0$  and the elements of  $B$  corresponding to lagged  $Y_{2t}$  are equal to zero. Finally, I refer to the general model given by (18) as the ‘censored and kinked SVAR’ (CKSVAR).

Define the  $k \times k$  square matrices

$$\bar{A} := \begin{pmatrix} A_{11} & A_{12} + A_{12}^* \\ A_{21} & A_{22} + A_{22}^* \end{pmatrix}, \quad \text{and} \quad A^* := \begin{pmatrix} A_{11} & A_{12}^* \\ A_{21} & A_{22}^* \end{pmatrix}. \quad (19)$$

$\bar{A}$  determines the impact effects of structural shocks during periods when the constraint does not bind.

$A^*$  does the same for periods when the constraint binds.

To analyze the CKSVAR, we first need to establish existence and uniqueness of the reduced form. This is done in the following proposition.

**Proposition 1** *The model given in eq. (18) is coherent and complete (i.e., it has a unique solution) if and only if*

$$\kappa := \frac{\bar{A}_{22} - A_{21}A_{11}^{-1}\bar{A}_{12}}{A_{22}^* - A_{21}A_{11}^{-1}A_{12}^*} > 0. \quad (20)$$

Note that (20) does not depend on the coefficients on the lags (whether latent or observed), so it is exactly the same as in a static SEM. This condition is useful for inference, e.g., when constructing confidence intervals or posteriors, because it restricts the range of admissible values for the structural

parameters. It can also be checked empirically when the structural parameters are point-identified.

If condition (20) is satisfied, there exists a reduced-form representation of the CKSVAR model (18). For convenience of notation, define the indicator (dummy variable) that takes the value one if the constraint binds and zero otherwise:

$$D_t = 1_{\{Y_{2t}=b\}}. \quad (21)$$

**Proposition 2** *If (20) holds, and for any initial values  $Y_{-s}, Y_{2,-s}^*$ ,  $s = 0, \dots, p-1$ , the reduced-form representation of (18) for  $t \geq 1$  is given by*

$$Y_{1t} = \bar{C}_1 X_t + \bar{C}_1^* \bar{X}_t^* + u_{1t} - \tilde{\beta} D_t \left( \bar{C}_2 X_t + \bar{C}_2^* \bar{X}_t^* + u_{2t} - b \right) \quad (22)$$

$$Y_{2t} = \max \left( \bar{Y}_{2t}^*, b \right), \quad (23)$$

$$\bar{Y}_{2t}^* = \bar{C}_2 X_t + \bar{C}_2^* \bar{X}_t^* + u_{2t}, \quad (24)$$

$$Y_{2t}^* = (1 - D_t) \bar{Y}_{2t}^* + D_t \left( \kappa \bar{Y}_{2t}^* + (1 - \kappa) b \right), \quad (25)$$

where  $u_t = (u'_{1t}, u'_{2t})' = \bar{A}^{-1} \varepsilon_t$ ,  $\bar{C}^* = (\bar{C}_1^*, \bar{C}_2^*)' = \kappa \bar{A}^{-1} B^*$ ,  $\bar{X}_t^* = (\bar{x}_{t-1}, \dots, \bar{x}_{t-p})'$ ,  $\bar{x}_t = \min \left( \bar{Y}_{2t}^* - b, 0 \right)$ ,  $\bar{x}_{-s} = \kappa^{-1} \min \left( Y_{2,-s}^* - b, 0 \right)$ ,  $s = 0, \dots, p-1$ ,

$$\tilde{\beta} = (A_{11} - A_{12}^* A_{22}^{*-1} A_{21})^{-1} (A_{12}^* A_{22}^{*-1} A_{22} - A_{12}), \quad (26)$$

$\kappa$  is defined in (20) and the matrices  $\bar{C}_1, \bar{C}_2$ , are given in eq. (48) in the Appendix.

Note that the “reduced-form” latent process  $\bar{Y}_{2t}^*$  is, in general, different from the “structural” shadow rate  $Y_{2t}^*$  defined by (25). They coincide only when  $\kappa = 1$ . This holds, for example, in the CSVAR model.

Equation (23) combined with (24) is a familiar dynamic Tobit regression model with the added complexity of latent lags included as regressors whenever  $\bar{C}_2^* \neq 0$ . Likelihood estimation of the univariate version of this model was studied by Lee (1999). The  $k-1$  equations (22) are ‘incidentally kinked’ dynamic regressions, that I have not seen analyzed before.

### 3.1 Identification

#### 3.1.1 Identification of reduced-form parameters

Let  $\psi$  denote the parameters that characterize the reduced form (22)-(23):  $\tilde{\beta}, \bar{C}, \bar{C}^*$  and  $\Omega = \text{var}(u_t)$ . It is useful to decompose  $\psi$  into  $\psi_2 = (\bar{C}_2, \bar{C}_2^*, \tau)'$ , where  $\tau = \sqrt{\text{var}(u_{2t})}$ , and  $\psi_1 = \left( \text{vec}(\bar{C}_1)', \text{vec}(\bar{C}_1^*)', \tilde{\beta}', \delta', \text{vech}(\Omega_{1.2}) \right)'$ , where  $\delta = \Omega_{12}/\tau^2$ ,  $\Omega_{1.2} = \Omega_{11} - \delta \delta' \tau^2$ , and  $\Omega_{ij} = \text{cov}(u_{it}, u_{jt})$ .

Equation (23) is the dynamic Tobit regression model studied by Lee (1999). So, its parameters,

$\psi_2$ , are generically identified provided that the regressors are not perfectly collinear. This requires that  $0 < \Pr(D_t = 1) < 1$ .

Given  $\psi_2$ , the identification of the remaining parameters,  $\psi_1$ , can be characterized using a control function approach. Consider the  $k - 1$  regression equations

$$E(Y_{1t}|Y_{2t}, X_t, \bar{X}_t^*) = \bar{C}_1 X_t + \bar{C}_1^* \bar{X}_t^* + \tilde{\beta} Z_{1t} + \delta Z_{2t}, \quad (27)$$

where

$$Z_{1t} = D_t \left( b - \bar{C}_2 X_t - \bar{C}_2^* \bar{X}_t^* - \frac{\tau \phi(a_t)}{\Phi(a_t)} \right), \quad (28)$$

$$Z_{2t} = (1 - D_t) \left( Y_{2t} - \bar{C}_2 X_t - \bar{C}_2^* \bar{X}_t^* \right) + D_t \frac{\tau \phi(a_t)}{\Phi(a_t)}, \quad (29)$$

$a_t = \left( \frac{b - \bar{C}_2 X_t - \bar{C}_2^* \bar{X}_t^*}{\tau} \right)$ , and  $\phi(\cdot)$ ,  $\Phi(\cdot)$  are the standard Normal density and distribution functions, respectively. When  $\bar{C}^*$  is different from zero, regressors  $\bar{X}_t^*$ ,  $Z_{1t}$ , and  $Z_{2t}$  in (27) are unobserved, so we need to replace them with their expectations conditional on  $Y_{2t}, Y_{t-1}, \dots, Y_1$ . Then, the regressors on the right-hand side of (27) become  $\mathbf{X}_t := (X_t', \bar{X}_{t|t}^{*'}, Z_{1t|t}, Z_{2t|t})'$ , where  $h_{t|t} := E(h(\bar{X}_t^*) | Y_{2t}, Y_{t-1}, \dots, Y_1)$  for any function  $h(\cdot)$  whose expectation exists.<sup>15</sup> The coefficients  $\bar{C}_1, \bar{C}_1^*, \tilde{\beta}$ , and  $\delta$  are generically identified if the regressors  $\mathbf{X}_t$  are not perfectly collinear.

### 3.1.2 Identification of structural parameters

From the order condition, we can easily establish that there are not enough restrictions to identify all the structural parameters in the CKSVAR (18). Let  $k_0 = \dim(X_{0t})$  denote the number of predetermined variables other than the own lags of  $Y_t$ . For example, in a standard VAR without deterministic trends, we have  $X_{0t} = 1$ , so  $k_0 = 1$ . The number of reduced-form parameters  $\psi$  is  $k_0 k + k^2 p$  (in  $\bar{C}$ ) plus  $k p$  (in  $\bar{C}^*$ ) plus  $k - 1$  (in  $\tilde{\beta}$ ) plus  $k(k + 1)/2$  (in  $\Omega$ ). The number of structural parameters in (18) is  $k_0 k + k^2 p$  (in  $B$ ) plus  $k p$  (in  $B^*$ ) plus  $k^2$  (in  $\bar{A}$ ) plus  $k$  (in  $A_{12}^*$  and  $A_{22}$ ). So, the CKSVAR is underidentified by  $k(k - 1)/2 + 1$  restrictions. Nevertheless, I will show that the impulse responses to  $\varepsilon_{2t}$  are identified. Specifically, they are point-identified when  $A_{12}^* = 0$ , and partially identified when  $A_{12}^* \neq 0$  but  $A_{12}^*$  and  $A_{12}$  have the same sign, analogous to the bounds given in equation (15) in the previous section.

Because the CKSVAR is nonlinear, IRFs are obviously state-dependent, and there are many ways one can define them, see Koop et al. (1996).<sup>16</sup> The IRF to  $\varepsilon_{2t}$ , according to any of the definitions proposed in the literature, is identified if the reduced-form errors  $u_t$  can be expressed as a known

<sup>15</sup>In the KSVAR model, we have  $\bar{C}_1^* = 0$  and  $\bar{C}_2^* = 0$ , so  $\bar{X}_t^*$  drops out of (27), and the regressors  $Z_{1t}, Z_{2t}$  are observed, so  $Z_{j|t} = Z_{jt}$ ,  $j = 1, 2$ .

<sup>16</sup>For the empirical analysis below, I will use the conditional “generalized impulse response function” defined in (Koop et al., 1996, eq. (3)), see (38).

function of  $\varepsilon_{2t}$  and a process that is orthogonal to it, i.e.,  $u_t = g(\varepsilon_{2t}, e_t)$ , where  $e_t$  is independent of  $\varepsilon_{2t}$ . From Proposition 2, it follows that the function  $g$  is linear, and more specifically,

$$u_{1t} = (I_{k-1} - \bar{\beta}\bar{\gamma})^{-1} (\bar{\varepsilon}_{1t} + \bar{\beta}\bar{\varepsilon}_{2t}), \text{ and} \quad (30)$$

$$u_{2t} = (1 - \bar{\gamma}\bar{\beta})^{-1} (\bar{\varepsilon}_{2t} + \bar{\gamma}\bar{\varepsilon}_{1t}), \quad (31)$$

where

$$\begin{aligned} \bar{\beta} &:= -A_{11}^{-1}\bar{A}_{12}, \quad \bar{\gamma} := -\bar{A}_{22}^{-1}A_{21}, \\ \bar{\varepsilon}_{1t} &:= A_{11}^{-1}\varepsilon_{1t}, \quad \bar{\varepsilon}_{2t} := \bar{A}_{22}^{-1}\varepsilon_{2t}, \end{aligned} \quad (32)$$

and  $\bar{A}_{22} = A_{22}^* + A_{22}$ , defined in (19). Note that  $\bar{\beta}$  can be interpreted as the response of  $Y_{1t}$  to a shock that increases  $Y_{2t}$  by one unit, and  $\bar{\gamma}$  are the contemporaneous reaction function coefficients of  $Y_{2t}$  to  $Y_{1t}$  when  $Y_{2t} > b$  (unconstrained regime). The shock vector  $\bar{\varepsilon}_{1t}$  is not structural but it is orthogonal to  $\varepsilon_{2t}$ , so it plays the role of  $e_t$  in  $u_t = g(\varepsilon_{2t}, e_t)$ . Hence, the IRF is identified if and only if  $\bar{\beta}$ ,  $\bar{\gamma}$ , and  $\bar{A}_{22}$  are identified.

The following proposition shows identification when  $A_{12}^* = 0$ .

**Proposition 3** *When  $A_{12}^* = 0$  and the coherency condition (20) holds, the parameters in (30)-(31) are identified by the equations  $\bar{\beta} = \tilde{\beta}$ ,*

$$\bar{\gamma} = \left( \Omega'_{12} - \Omega_{22}\bar{\beta}' \right) \left( \Omega_{11} - \Omega_{12}\bar{\beta}' \right)^{-1}, \text{ and} \quad (33)$$

$$\bar{A}_{22}^{-1} = \sqrt{(-\bar{\gamma}, 1) \Omega (-\bar{\gamma}, 1)'}. \quad (34)$$

**Remarks** 1.  $\bar{\beta} = \tilde{\beta}$  follows immediately from the definition (26) with  $A_{12}^* = 0$ . Equations (33) and (34) hold without the restriction  $A_{12}^* = 0$ . They follow from the orthogonality of the shocks  $\varepsilon_{2t}$  and  $\bar{\varepsilon}_{1t}$ .

2. An instrumental variables interpretation of this identification result is as follows. Define the instrument

$$Z_t := Y_{1t} - \tilde{\beta}Y_{2t} = A_{11}^{-1}B_1X_t + A_{11}^{-1}B_1^*X_t^* + A_{11}^{-1}\varepsilon_{1t},$$

where the second equality holds when  $A_{12}^* = 0$ . The orthogonality of the errors  $E(\varepsilon_{1t}\varepsilon_{2t}) = 0$  implies  $E(Z_t\varepsilon_{2t}) = 0$ . So,  $Z_t$  are valid  $k-1$  instruments for the  $k-1$  endogenous regressors  $Y_{1t}$  in the structural equation of  $Y_{2t} = \max(Y_{2t}^*, b)$ , where  $Y_{2t}^*$  is given by (17). Normalizing (17) in terms of  $Y_{2t}^*$  yields the structural equation in the more familiar form of a policy rule:

$$Y_{2t} = \max(\bar{\gamma}Y_{1t} + \bar{B}_2X_t + \bar{B}_2^*X_t^* + \bar{\varepsilon}_{2t}, b), \quad (35)$$

where  $\bar{B}_2 = \bar{A}_{22}^{-1} B_2$ ,  $\bar{B}_2^* = \bar{A}_{22}^{-1} B_2^*$ . Since  $A_{11}^{-1}$  is non-singular, the coefficient matrix of  $Z_t$  in the ‘first-stage’ regressions of  $Y_{1t}$  is nonsingular, so the coefficients of (35) are generically identified by the rank condition. An alternative to the Tobit IV regression model (35) is the indirect Tobit regression approach used in the static SEM by Blundell and Smith (1994). Equation (35) can be written as the dynamic Tobit regression

$$Y_{2t} = \max \left( \tilde{\gamma} Z_t + \tilde{B}_2 X_t + \tilde{B}_2^* X_t^* + \tilde{\varepsilon}_{2t}, b \right), \quad (36)$$

where  $\tilde{\gamma} = (1 - \bar{\gamma}\bar{\beta})^{-1} \bar{\gamma}$ ,  $\tilde{B}_2 = (1 - \bar{\gamma}\bar{\beta})^{-1} \bar{B}_2$ ,  $\tilde{B}_2^* = (1 - \bar{\gamma}\bar{\beta})^{-1} \bar{B}_2^*$  and  $\tilde{\varepsilon}_{2t} = (1 - \bar{\gamma}\bar{\beta})^{-1} \bar{\varepsilon}_{2t}$ . Note that the coherency condition (20) becomes  $\kappa = \frac{\bar{A}_{22}}{A_{22}^*} (1 - \bar{\gamma}\bar{\beta}) > 0$ , so  $1 - \bar{\gamma}\bar{\beta} \neq 0$ , which guarantees the existence of the representation (36). Given  $\bar{\beta} = \tilde{\beta}$ , the structural parameter  $\bar{\gamma}$  can then be obtained as  $\bar{\gamma} = \tilde{\gamma} \left( I_{k-1} + \tilde{\beta} \tilde{\gamma} \right)^{-1}$ , and similarly for the remaining structural parameters in (35).

3. The parameter  $A_{22}$  allows the reaction function of  $Y_{2t}^*$  to differ across the two regimes. The special case  $A_{22} = 0$  thus corresponds to the restriction that the reaction function remains the same across regimes. The parameters  $A_{22}$  and  $A_{22}^*$  are not separately identified. Hence,  $A_{22}^{*-1}$ , the scale of the response to the shock  $\varepsilon_{2t}$  during periods when  $Y_{2t} = b$ , is not identified.<sup>17</sup> Similarly,  $\kappa = \frac{\bar{A}_{22}}{A_{22}^*} (1 - \bar{\gamma}\bar{\beta})$  is not identified, and therefore, neither is the structural shadow value  $Y_{2t}^*$  in eq. (25). Identification of these requires an additional restriction on  $A_{22}$ , e.g.,  $A_{22} = 0$ . Turning this discussion around, we see that a change in the reaction function across regimes does not destroy the point identification of the effects of policy during the unconstrained regime, since the latter only requires  $\bar{\beta}, \bar{\gamma}$  and  $\bar{A}_{22}$ , not  $A_{22}^*$  or  $\kappa$ .

Next, we turn to the case  $A_{12}^* \neq 0$ , and derive identification under restrictions on the sign and magnitude of  $A_{12}^*$  relative to  $A_{12}$  and  $A_{22}^*$  relative to  $A_{22}$ . The first restriction is motivated by a generalization of the discussion on the SEM model in Subsection 2.1.2. Specifically, if  $\bar{A}_{12} = A_{12} + A_{12}^*$  measures the effect of conventional policy (operating in the unconstrained regime) and  $A_{12}^*$  measures the effect of unconventional policy (operating in the constrained regime), then the assumption that  $A_{12}$  and  $A_{12}^*$  have the same sign means that unconventional policy effects are neither in the opposite direction nor larger in absolute value than conventional policy effects. In other words, unconventional policy is neither counterproductive nor over-productive relative to conventional policy. This can be characterized by the specification  $A_{12}^* = \Lambda \bar{A}_{12}$  and  $A_{12} = (I_{k-1} - \Lambda) \bar{A}_{12}$ , where  $\Lambda = \text{diag}(\lambda_j)$ ,  $\lambda_j \in [0, 1]$  for  $j = 1, \dots, k-1$ . I further impose the restriction that  $\lambda_j = \lambda$  for all  $j$ , so that  $A_{12}^* = \lambda \bar{A}_{12}$  and  $A_{12} = (1 - \lambda) \bar{A}_{12}$  with  $\lambda \in [0, 1]$ . This, in turn, means that  $Y_{2t}$  and  $Y_{2t}^*$  enter each of the first  $k-1$  structural equations for  $Y_{1t}$  only via the common linear combination  $\lambda Y_{2t}^* + (1 - \lambda) Y_{2t}$ , which can be interpreted as a measure of the effective policy stance.

<sup>17</sup>This is akin to the well-known property of a probit model that the scale of the distribution of the latent process is not identifiable.

We also need to consider the impact of  $A_{22}$  on identification. The parameter  $\zeta = \bar{A}_{22}/A_{22}^*$  gives the ratio of the standard deviation of the monetary policy shock in the constrained relative to the unconstrained regime. It is also the ratio of the reaction function coefficients in the two regimes, e.g.,  $A_{22}^{*-1}A_{21}$  versus  $\bar{A}_{22}^{-1}A_{21}$ . I will impose  $\zeta > 0$ , so that the sign of the policy shock does not change across regimes. With the above reparametrization and the definitions in (32), the identified coefficient  $\tilde{\beta}$  in (8) can be written as

$$\tilde{\beta} = (1 - \xi) (I - \xi \bar{\beta} \bar{\gamma})^{-1} \bar{\beta}, \quad \xi := \lambda \zeta. \quad (37)$$

Similarly, given  $\zeta > 0$ , the coherency condition (20) reduces to  $(1 - \bar{\gamma} \bar{\beta}) (1 - \xi \bar{\gamma} \bar{\beta}) > 0$ . Notice that the parameters  $\lambda, \zeta$  only appear multiplicatively, so it suffices to consider them together as  $\xi = \lambda \zeta$ . Once  $\bar{\beta}$  is known, the remaining structural parameters needed to obtain the IRF to  $\varepsilon_{2t}$  are  $\bar{\gamma}$  and  $\bar{A}_{22}$ , and they are obtained from Proposition 3. So, the identified set can be characterized by varying  $\xi$  over its admissible range. Without further restrictions on  $\zeta$ , the admissible range is obviously  $\xi \geq 0$ . If we further assume that  $\zeta \leq 1$ , i.e., that the slope of the reaction function coefficients is no steeper in the constrained regime than in the unconstrained regime, then  $\xi \in [0, 1]$ , and so partial identification proceeds exactly along the lines of the SEM in the previous section where  $\lambda$  played the role of  $\xi$ . In the case  $k = 2$ , the bounds derived in eq. (15) apply, with  $\beta = \bar{\beta}$  in the notation of the present section. However, when  $k > 2$ , it is difficult to obtain a simple analytical characterization of the identified set for  $\bar{\beta}$ . In any case, we will typically wish to obtain the identified set for functions of the structural parameters, such as the IRF. This can be done numerically by searching over a fine discretization of the admissible range for  $\xi$ . An algorithm for doing this is provided in Appendix B.2.

### 3.2 Estimation

Estimation of the KSVAR is straightforward, since the likelihood is analytically available under Gaussian errors, see equation (62) in Appendix B.1. The key is that latent lags do not appear on the right-hand side of the model. Estimation of the CKSVAR is more involved because the presence of latent lags means that the likelihood function is not analytically available. It can be approximated using particle filtering, such as the sequential importance sampler (SIS) proposed by Lee (1999) for a univariate dynamic Tobit model. The SIS has the attractive feature that it is smooth, so the likelihood can be numerically differentiated and maximized with derivatives-based methods. However, the SIS can potentially suffer from sample degeneracy, see Herbst and Schorfheide (2015). I therefore consider also a fully adapted particle filter (FAPF) that uses resampling to address the sample degeneracy problem. The method can be found in Malik and Pitt (2011) and is a special case of the auxiliary particle filter of Pitt and Shephard (1999). The disadvantage of FAPF is that the resampling step makes it discontinuous, so the likelihood cannot be maximized using derivatives-based algorithms, nor



can we compute standard errors using numerical differentiation. We can still maximize the likelihood using simulated annealing and use the Likelihood Ratio (LR) test for inference. Given the aforementioned computational challenges, it is possibly more practical to implement FAPF using Bayes than Maximum Likelihood (ML).

The description of the likelihood and the two filtering algorithms is given in Appendix B. In addition, one can consider (simulated) method of moments, and sequential estimation motivated by the constructive identification proof in Section 3. These are described in the Supplementary Appendix.

When the data is stationary and ergodic, and subject to some additional regularity conditions given in Newey and McFadden (1994), the ML estimator can be shown to be consistent and asymptotically Normal and the LR statistic asymptotically  $\chi^2$  with degrees of freedom equal to the number of restrictions. I will not discuss primitive conditions for these results, but I note that they are somewhat weaker than for a standard linear VAR. This is because the lower bound on one of the variables means that autoregressive roots that are lower than  $-1$  may not violate stationarity, see, e.g., (de Jong and Herrera, 2011, p. 230). Instead, I report Monte Carlo simulation results on the finite-sample properties of ML estimators and LR tests in Appendix C. They show that the Normal distribution provides a very good approximation to the finite-sample distribution of the ML estimators. I find some finite-sample size distortion in the LR tests of various restrictions on the CKSVAR, but this can be addressed effectively with a parametric bootstrap, as shown in the Appendix.

One interesting observation from the simulations is that the LR test of the CSVAR restrictions against the CKSVAR appears to be less powerful than the corresponding test of the KSVAR restrictions against the CKSVAR. Thus, we expect to be able to detect deviations from KSVAR more easily than deviations from CSVAR. In other words, finding evidence against the hypothesis that unconventional policies are fully effective (CSVAR) may be harder than finding evidence against the opposite hypothesis that they are completely ineffective (KSVAR).

## 4 Application

I use the three-equation SVAR of Stock and Watson (2001), consisting of inflation, the unemployment rate and the Federal Funds rate to provide a simple empirical illustration of the methodology developed in this paper. As discussed in Stock and Watson (2001), this model is far too limited to provide credible identification of structural shocks, so the results in this section are meant as an illustration of the new methods.

The data are quarterly and are constructed exactly as in Stock and Watson (2001).<sup>18</sup> The variables are plotted in Figure 1 over the extended sample 1960q1 to 2018q2. I will consider all periods in which the Fed funds rate was below 20 basis points to be on the ZLB. This includes 28 quarters, or 11% of

---

<sup>18</sup>The inflation data are computed as  $\pi_t = 400\ln(P_t/P_{t-1})$ , where  $P_t$  is the implicit GDP deflator and  $u_t$  is the civilian unemployment rate. Quarterly data on  $u_t$  and  $i_t$  are formed by taking quarterly averages of their monthly values.

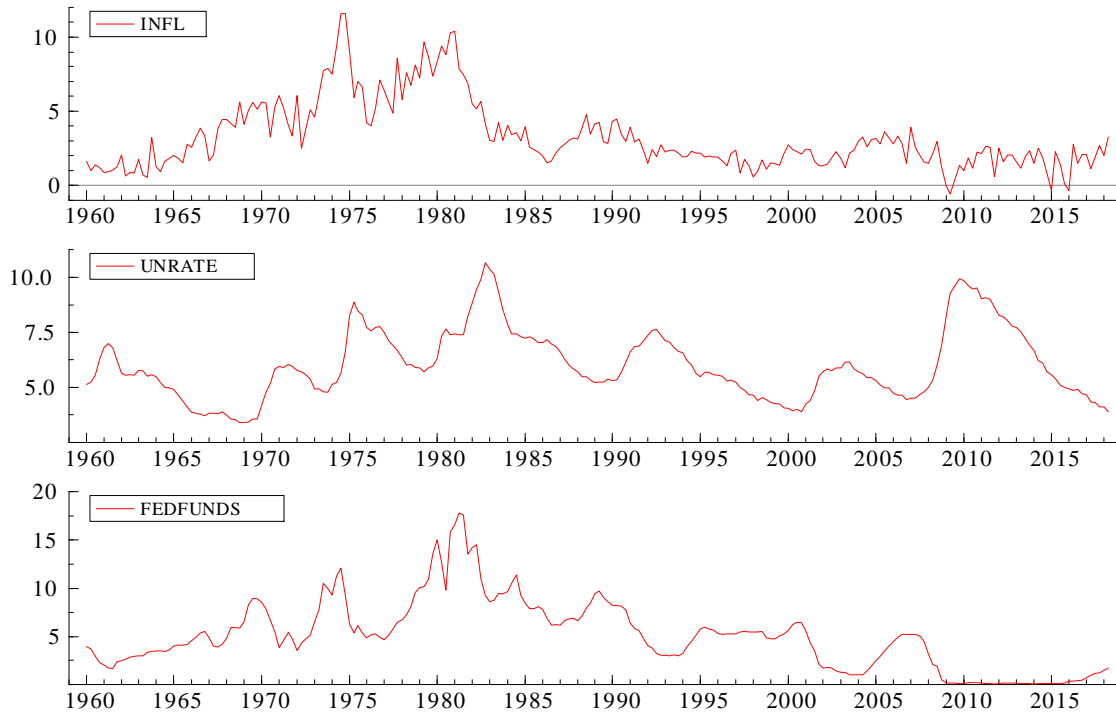


Figure 1: Data used in Stock and Watson (2001) over the extended sample 1960q1 to 2018q2.

Model	log lik	(FAPF)	# restr.	LR stat.	Asym. p-val.	Boot. p-val.
CKSVAR(4)	-81.64	-81.94				
KSVAR(4)	-97.05	-	12	30.82	0.002	0.011
CSVAR(4)	-94.86	-94.87	14	26.43	0.023	0.117

Sample: 1960q1-2018q2 (234 obs, 11% at ZLB)

Table 1: Maximized log-likelihood of various SVAR models in inflation, unemployment and Federal funds rate. CKSVAR corresponds to the unrestricted specification (22)-(23); KSVAR excludes latent lags; CSVAR is a purely censored model. CKSVAR and CSVAR likelihoods computed using sequential importance sampling with 1000 particles (alternative estimates based on Fully Adapted Particle Filtering with resampling are shown in parentheses). Asymptotic p-values from  $\chi^2_q$ ,  $q$  = number of restrictions. Bootstrap p-values from parametric bootstrap with 999 replications.

the sample.

#### 4.1 Tests of efficacy of unconventional policy

I estimate three specifications of the SVAR(4) with the ZLB: the unrestricted CKSVAR specification, as well as the restricted KSVAR and CSVAR specifications. The maximum log-likelihood for each model is reported in Table 1, computed using the SIS algorithm in the case of CKSVAR and CSVAR, with 1000 particles. The accuracy of the SIS algorithm was gauged by comparing the log-likelihood to the one obtained using the resampling FAPF algorithm. In both CKSVAR and CSVAR the difference is very small. The results are also very similar when we increase the number of particles to 10000. Finally, the table reports the likelihood ratio tests of KSVAR and CSVAR against CKSVAR using both asymptotic and parametric bootstrap p-values.

The KSVAR imposes the restriction that no latent lags (i.e., lags of the shadow rate) should appear on the right hand side of the model, i.e.,  $B^* = 0$  in (18) or  $\bar{C}_1^* = 0$  and  $\bar{C}_2^* = 0$  in (22) and (23). This amounts to 12 exclusion restrictions on the CKSVAR(4), four restrictions in each of the three equations. This is necessary (but not sufficient) for the hypothesis that unconventional policy is completely ineffective at all horizons. It is necessary because  $\bar{C}^* = (\bar{C}_1^{*'}, \bar{C}_2^{*'})' \neq 0$  would imply that unconventional policy has at least a lagged effect on  $Y_t$ .  $\bar{C}^* = 0$  is not sufficient to infer that unconventional policy is completely ineffective because it may still have a contemporaneous effect on  $Y_{1t}$  if  $A_{12}^* \neq 0$ , and the latter is not point-identified. The result of the test in Table 1 shows that lags of the shadow rate are statistically significant at the 5% level, rejecting the null hypothesis that unconventional policy has no effect.

The CSVAR model imposes the restriction that only the coefficients on the lags of the shadow rate (which is equal to the actual rate above the ZLB) are different from zero in the model, i.e., the elements of  $B$  corresponding to lags of  $Y_{2t}$  in (18) are all zero, or equivalently, the elements of  $\bar{C}$  corresponding to lags of  $Y_{2t}$  in (22) and (23) are all equal to  $\bar{C}^*$ . In addition, it imposes the restriction that  $\tilde{\beta} = 0$  in (22), i.e., no kink in the reduced-form equations for inflation and unemployment across regimes, yielding 14 restrictions in total. This is necessary for the hypothesis that the ZLB is empirically irrelevant for policy in that it does not limit what monetary policy can achieve. The evidence against this hypothesis is not as strong as in the case of the KSVAR. The asymptotic p-value is 0.023, indicating rejection of the null hypothesis that unconventional policy is as effective as conventional policy at the 5% level, but the bootstrap p-value is 0.117. Note that this difference could also be due to fact that the test of the CSVAR restrictions may be less powerful than the test of the KSVAR restrictions, as indicated by the simulations reported in the previous section. Thus, I would cautiously conclude that the evidence on the empirical relevance of the ZLB is mixed. Further evidence on the efficacy of unconventional policy will also be provided in the next subsection.

## 4.2 IRFs

Based on the evidence reported in the previous section, I estimate the IRFs associated with the monetary policy shock using the unrestricted CKSVAR specification, and compare them to recursive IRFs from the CSVAR specification that place the Federal funds rate last in the causal ordering. From the identification results in Section 3, the CKSVAR point-identifies the nonrecursive IRFs only under the assumption that the shadow rate has no contemporaneous effect of  $Y_{1t}$ , i.e.,  $A_{12}^* = 0$  in (16). Note that, due to the nonlinearity of the model, the IRFs are state-dependent. I use the following definition of the IRF from Koop et al. (1996):

$$IRF_{h,t}(\varsigma, X_t, \bar{X}_t^*) = E(Y_{t+h} | \varepsilon_{2t} = \varsigma, X_t, \bar{X}_t^*) - E(Y_{t+h} | \varepsilon_{2t} = 0, X_t, \bar{X}_t^*). \quad (38)$$

If  $t$  is such that  $\bar{Y}_{2,t-s}^* = Y_{2,t-s} > b$  for all  $s = 1, \dots, p$ , then  $\bar{X}_t^* = 0$ ,<sup>19</sup> and hence  $IRF_{h,t}(\varsigma, X_t, 0)$  is observed. However, if  $t$  is such that any element of  $\bar{X}_t^*$  is unobserved, we could either evaluate  $IRF_{h,t}(\varsigma, X_t, \cdot)$  at an estimated value of  $\bar{X}_t^*$  conditional on the observed data, or we can estimate  $IRF_{h,t}(\varsigma, X_t, \bar{X}_t^*)$  conditional on the observed data, see the Supplementary Appendix for more details. When we estimate the IRF from a CKSVAR(4) at the end of our sample,  $t = 2018q3$ ,  $\bar{X}_t^* = 0$  because the Federal funds rate was above the ZLB over the previous four quarters, so filtering is unnecessary.

Figure 2 reports the nonrecursive IRFs to a 25 basis points monetary policy shock from the CKSVAR under the assumption that  $\lambda = 0$  (unconventional policy has no contemporaneous effect) and two different estimates of recursive IRFs using the identification scheme in Stock and Watson (2001). The first estimate is a nonlinear IRF that is obtained from the CSVAR specification. The second is a “naive” OLS estimate of the linear IRF in a SVAR with interest rates placed last, ignoring the ZLB constraint (a direct application of the method in Stock and Watson (2001) to the present sample). The figure also reports 90% bootstrap error bands for the nonrecursive IRFs.

In the nonrecursive IRF, the response of inflation to a monetary tightening is negative on impact, albeit very small, and, with the exception of the first quarter when it is positive, it stays negative throughout the horizon. Hence, the incidence of a price puzzle is mitigated relative to the recursive IRFs, according to which inflation rises for up to 6 quarters after a monetary tightening (9 quarters in the OLS case). Note, however, that the error bands are so wide that they cover (pointwise) most of the recursive IRF, though less so for the OLS one. Turning to the unemployment response, we see that the nonrecursive IRF starts significantly positive on impact (no transmission lag) and peaks much earlier (after 4 quarters) than the recursive IRF (10 quarters). In this case, the recursive IRF is outside the error bands for several quarters (more so for the naive OLS IRF). Finally, the response of the Federal funds rate to the monetary tightening is less than one on impact and generally significantly lower than the recursive IRFs. This is both due to the contemporaneous feedback from inflation and unemployment, as well as the fact that there is a considerable probability of returning to the ZLB, which mitigates the impact of monetary tightening.

Next, I turn to the identified sets of the IRFs that arise when I relax the restriction that unconventional policy is ineffective, i.e.,  $\lambda$  can be greater than zero. I consider the range of  $\xi = \lambda\zeta \in [0, 1]$ , recalling that  $\lambda$  measures the efficacy of unconventional policy and  $\zeta$  measures the ratio of the reaction function coefficients and shock volatilities in the constrained versus the unconstrained regimes. The left-hand-side graphs in Figure 3 report the identified sets without any other restrictions. The right-hand-side graphs derive the identified sets when I impose the additional sign restriction that the contemporaneous effect of the monetary policy shock to the Fed Funds rate should be nonnegative. The red shaded area gives the identified sets and the blue line with diamonds gives the point estimates when  $\lambda = 0$ . The latter are the same as the nonrecursive point estimates reported in Figure 2.

<sup>19</sup>Recall the definition  $\bar{X}_t^* = (\bar{x}_{t-1}, \dots, \bar{x}_{t-p})'$ ,  $\bar{x}_t = \min(\bar{Y}_{2t}^* - b, 0)$ , given in Proposition 2.

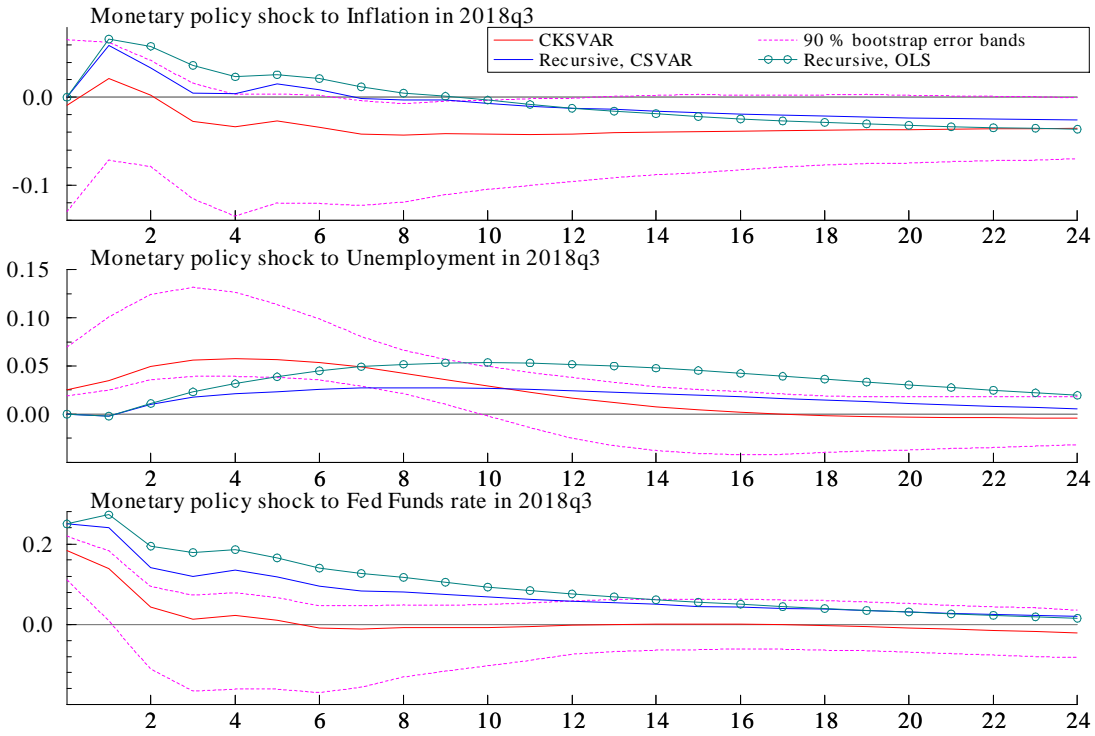


Figure 2: IRFs to a 25bp monetary policy shock from a three equation CKSVAR(4) estimated over the period 1960q1 to 2018q2. The solid red line corresponds to the nonrecursive identification from the ZLB under the assumption that unconventional policy is ineffective. The dashed blue line corresponds to the nonlinear recursive IRF, estimated with the CSVAR(4) under the restriction that the contemporaneous impact of Fed Funds on inflation and unemployment is zero. The green line with circles corresponds to the recursive IRF from a linear SVAR(4) estimated by OLS with Fed Funds ordered last.

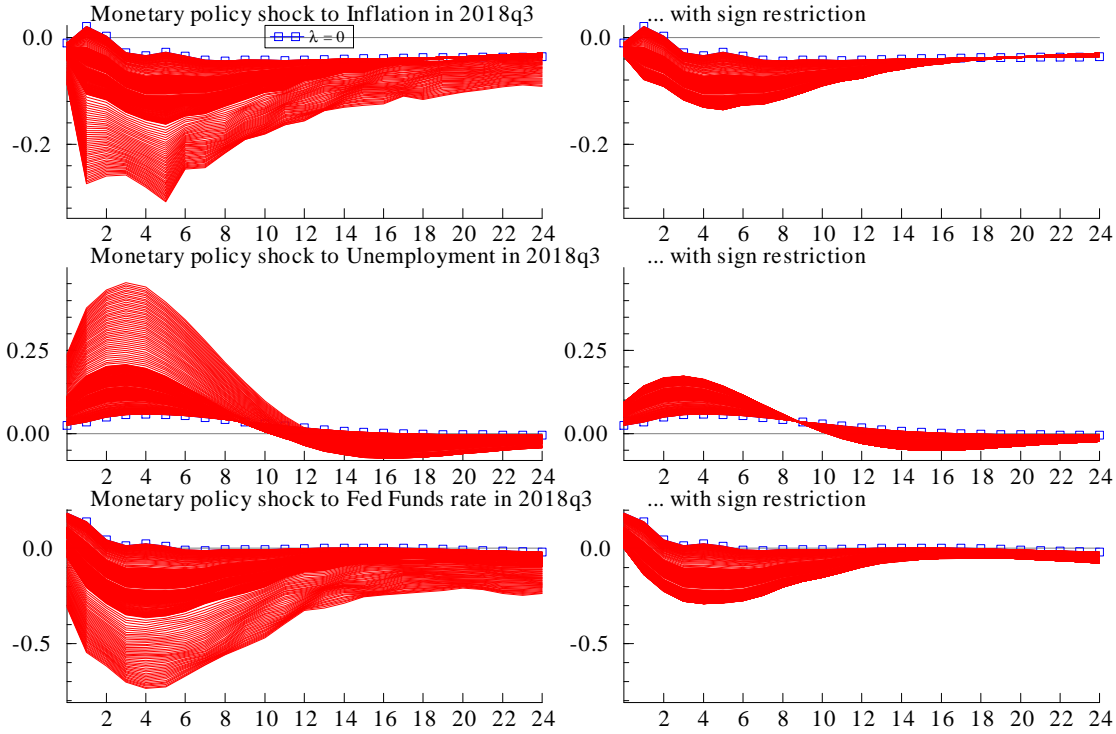


Figure 3: Identified sets of the IRFs to a 25bp monetary policy shock in a CKSVAR(4) in 2018q3. The red area denotes the identified set, the blue line with diamonds indicates the point identified IRF under the restriction that unconventional policy is ineffective ( $\lambda = 0$ ). The figures on the left provide the full identified set. The figures on the right impose the restriction that the response of the Fed funds rate be nonnegative.

We observe that the identified set for the IRF of inflation is bounded from above by the limiting case  $\lambda = 0$ . This is also true of the response of the Fed Funds rate. The case  $\lambda = 0$  provides a lower bound on the effect to unemployment only from 0 to 9 quarters. Even though the point estimate of the unemployment response under  $\lambda = 0$  remains positive over all horizons, the identified set includes negative values beyond 10 quarters ahead. We also notice that the identified sets are fairly large, albeit still informative. Interestingly, the identified IRF of the Fed Funds rate includes a range of negative values on impact. These values arise because for values of  $\xi > 0$ , there are generally two solutions for the structural VAR parameters  $\bar{\beta}, \bar{\gamma}$  in the equations (33), (37), with one of them inducing such strong responses of inflation and unemployment to the interest rate that the contemporaneous feedback in the policy rule would in fact revert the direct positive effect of the policy shock on the interest rate. If we impose the additional sign restriction that the contemporaneous impact of the policy shock to the Fed Funds rate must be non-negative, then those values are ruled out and the identified sets become considerably tighter. This is an example of how sign restrictions can lead to tighter partial identification of the IRF.

With an additional assumption on  $\zeta$ , the method can be used to obtain an estimate of the identified set for  $\lambda$ , the measure of the efficacy of unconventional policy. In particular, if we set  $\zeta = 1$ , i.e., the reaction function remains the same across the two regimes, then the identified set for  $\lambda$  is  $[0.0, 0.506]$ . In

other words, the identified set excludes values of the efficacy of policy beyond 51%, so that, roughly speaking, unconventional policy is at most 51% as effective as conventional one. Note that this estimate does not account for sampling uncertainty and relies crucially on the assumption that the reaction function remains the same across the two regimes. This assumption could be justified by arguing that there is no reason to believe that policy objectives may have shifted over the ZLB period, and that any desired policy stance was feasible over that period. The latter assumption may be questionable. For example, one can imagine that there may be financial and political constraints on the amount of quantitative easing policy makers could do, which may cause them to proceed more cautiously over the ZLB period than over regular times. Within the context of our model, this would be reflected as a flatter policy reaction function over the ZLB period than over the non-ZLB periods, i.e., it will correspond to  $\zeta < 1$ . To illustrate the implications of this for the identification of  $\lambda$ , suppose that  $\zeta = 1/2$ , i.e., the shadow rate reacts half as fast to shocks during the ZLB period than it does in the non-ZLB period. Then, the identified set for  $\lambda$  would include 1, i.e., the data would be consistent with the view that unconventional policy is fully effective. So, under this alternative assumption on  $\zeta$ , the reason we observed a subdued response to policy shocks over the ZLB period is because policy was less active over that period, and policy shocks were smaller, not because unconventional policy was partially ineffective.

As I discussed in the introduction, it is difficult to make further progress on this issue without further information or additional assumptions. The technical reason is that the scale of the latent regression over the censored sample is not identified, so additional information is required to untangle the structural parameters  $\lambda$  and  $\zeta$  from  $\xi = \lambda\zeta$ . One possibility would be to identify  $\lambda$  from the coefficients on the lags of  $Y_{2t}$  and  $Y_{2t}^*$  by imposing the (overidentifying) restriction that  $Y_{2t-j}$  and  $Y_{2t-j}^*$  appear in the model only via the linear combination  $Y_{2t-j}^{\text{eff}} := \lambda Y_{2t-j}^* + (1 - \lambda) Y_{2t-j}$  for all lags  $j = 1, \dots, p$ , where  $Y_{2t}^{\text{eff}}$  can be interpreted as the effective policy stance. Provided that the coefficients on the lags of  $Y_{2t}^*$  or  $Y_{2t}$  are not all zero, this restriction point identifies  $\lambda$ , and hence, partially identifies  $\zeta$  from  $\xi$ . One could obtain tighter bounds by using sign restrictions or obtain point identification by using conventional identification schemes.<sup>20</sup> For instance, one can identify  $\bar{\beta}$  directly using external instruments, as in Gertler and Karadi (2015), and hence point identify  $\xi$  from (37).

## 5 Conclusion

This paper has shown that the ZLB can be used constructively to identify the causal effects of monetary policy on the economy. Identification relies on the (in)efficacy of alternative (unconventional) policies. When unconventional policies are partially effective in mitigating the impact of the ZLB, the causal effects of monetary policy are only partially identified. A general method is proposed to estimate

---

<sup>20</sup>See Ikeda et al. (2020) for an application with sign restrictions.

SVARs subject to an occasionally binding constraint. The method can be used to test the efficacy of unconventional policy, modelled via a shadow rate. Application to a core three-equation SVAR with US data suggests that the ZLB is empirically relevant and unconventional policy is only partially effective.

## A Proofs

### A.1 Derivation of equation (15)

For any given  $\beta$ , the orthogonality of the errors implies that

$$\gamma = \frac{\omega_{12} - \omega_{22}\beta}{\omega_{11} - \omega_{12}\beta}, \quad (39)$$

see the proof of Proposition 3. Substituting this into (8) we have

$$\tilde{\beta} = g(\beta)\beta, \quad g(\beta) := \frac{1 - \lambda}{1 - \lambda \frac{\beta(\omega_{12} - \omega_{22}\beta)}{\omega_{11} - \omega_{12}\beta}}. \quad (40)$$

When  $\omega_{12} = 0$ , we have  $g(\beta) = \frac{1 - \lambda}{1 + \lambda\omega_{22}\beta^2/\omega_{11}} \in (0, 1)$  for all  $\lambda \in (0, 1)$ . Therefore, when  $\tilde{\beta} \neq 0$ , the sign of  $\beta$  is the same as that of  $\tilde{\beta}$  and its magnitude is lower, as stated in (15).

Next, consider  $\omega_{12} \neq 0$ . It is easily seen that  $g(0) = 1 - \lambda$  and  $\lim_{\beta \rightarrow \pm\infty} g(\beta) = 0$ . Moreover,

$$\frac{\partial g}{\partial \beta} = \lambda(1 - \lambda) \frac{\omega_{12}\omega_{22}\beta^2 - 2\omega_{11}\omega_{22}\beta + \omega_{11}\omega_{12}}{(\omega_{11} - \beta\omega_{12} - \beta\lambda\omega_{12} + \beta^2\lambda\omega_{22})^2}.$$

For  $\lambda \in (0, 1)$ , the above derivative function has zeros at

$$\omega_{12}\omega_{22}\beta^2 - 2\omega_{11}\omega_{22}\beta + \omega_{11}\omega_{12} = 0,$$

which occur at

$$\begin{aligned} \beta_1 &= \frac{\omega_{11}\omega_{22} + \sqrt{\omega_{11}\omega_{22}(\omega_{11}\omega_{22} - \omega_{12}^2)}}{\omega_{12}\omega_{22}} \\ \beta_2 &= \frac{\omega_{11}\omega_{22} - \sqrt{\omega_{11}\omega_{22}(\omega_{11}\omega_{22} - \omega_{12}^2)}}{\omega_{12}\omega_{22}}, \quad \text{if } \omega_{12} \neq 0. \end{aligned}$$

Now, because  $0 < (\omega_{11}\omega_{22} - \omega_{12}^2) < \omega_{11}\omega_{22}$  implies  $\sqrt{\omega_{11}\omega_{22}(\omega_{11}\omega_{22} - \omega_{12}^2)} < \omega_{11}\omega_{22}$ , we have  $\beta_i < 0$ ,  $i = 1, 2$ , when  $\omega_{12} < 0$  and  $\beta_i > 0$ ,  $i = 1, 2$ , when  $\omega_{12} > 0$ .

By symmetry, it suffices to consider only one of the two cases, e.g., the case  $\omega_{12} < 0$ . In this case,  $g'(\beta) = \frac{\partial g}{\partial \beta} < 0$  for all  $\beta > 0$  and, since  $g(0) = 1 - \lambda$  and  $g(\infty) = 0$ , it follows that  $g(\beta) \in (0, 1 - \lambda)$  for all  $\beta > 0$ . Thus, from (40) we see that  $\tilde{\beta} < 0$  cannot arise from  $\beta > 0$  when  $\omega_{12} < 0$ . In other words, observing  $\tilde{\beta} < 0$  must mean that  $\beta < 0$ . Moreover, since  $g'(\beta) < 0$  for all  $\beta > \beta_1$  and  $\beta_1 < 0$ , it must be that  $g(\beta) > 0$  for all  $\beta > \beta_1$ , and hence, also for  $\beta_1 < \beta \leq 0$ . At  $\beta < \beta_1$ ,  $g'(\beta) > 0$ , and since  $g'(\beta) < 0$  for all  $\beta < \beta_2 < \beta_1$ , and  $g(-\infty) = 0$ , it has to be that  $g(\beta)$  approaches zero from below as



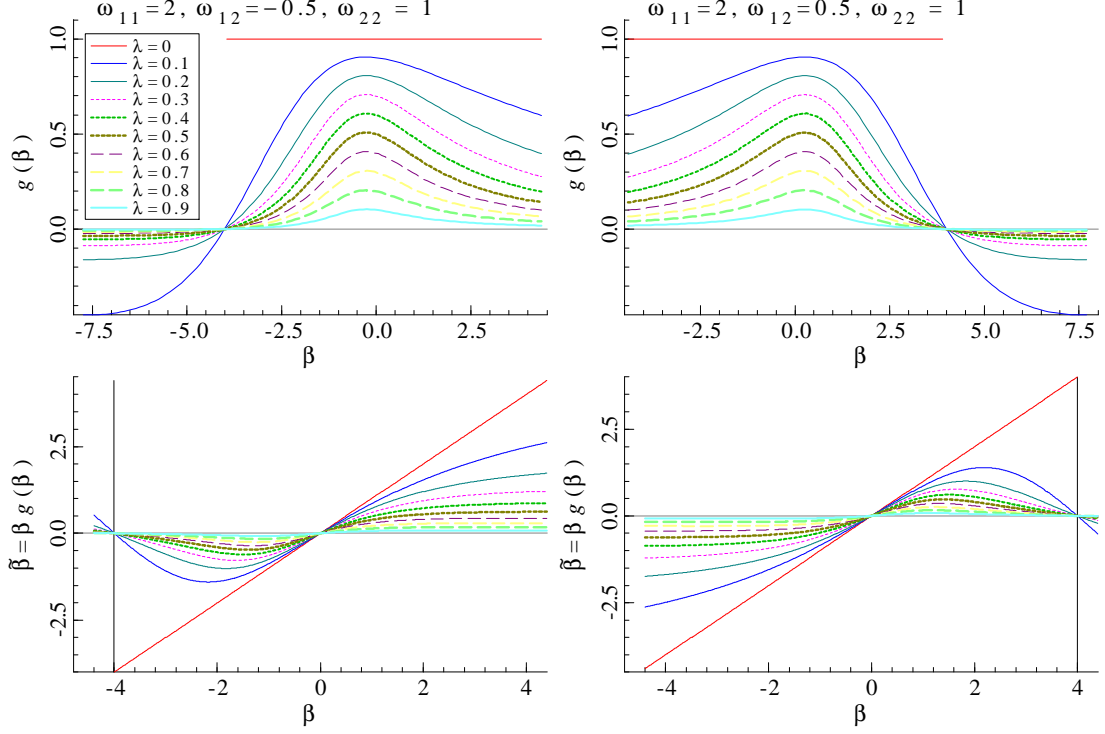


Figure 4: Plots function  $g(\beta) = \frac{1-\lambda}{1-\lambda \frac{\beta(\omega_{12}-\omega_{22}\beta)}{\omega_{11}-\omega_{12}\beta}}$  (top row) and  $\tilde{\beta} = g(\beta)\beta$  (bottom row), for two different parametrizations (in columns). The vertical line in the bottom row corresponds to  $\beta = \omega_{11}/\omega_{12}$ .

$\beta \rightarrow -\infty$ , and therefore,  $g(\beta)$  must cross zero at some  $\beta_0 \in (\beta_2, \beta_1)$ , and  $g(\beta) \geq 0$  for all  $\beta \in [\beta_0, 0]$ . Inspection of (40) shows that  $\beta_0 = \omega_{11}/\omega_{12} = 1/\gamma_0$ , which corresponds to  $\gamma = -\infty$  from (39). Since  $g(\beta) \in [0, 1-\lambda]$  for all  $\beta \in [\beta_0, 0]$ , and  $\lambda \in (0, 1)$ , it follows from (40) that  $|\tilde{\beta}| \leq |\beta|$ . In other words,  $\tilde{\beta}$  is attenuated relative to the true  $\beta$ .

Finally, we notice that there is a minimum value of  $\tilde{\beta}$  that one can observe under the restriction  $\lambda \in [0, 1]$  (at  $\lambda = 1$ ,  $\tilde{\beta} = 0$ ). Given the attenuation bias and the fact that  $\tilde{\beta} < 0$  if and only if  $\beta \in [\beta_0, 1]$ , the smallest value of  $\tilde{\beta}$  occurs when  $\lambda = 0$  and  $\beta = \omega_{11}/\omega_{12}$ , so  $\tilde{\beta}_{\min} = \omega_{11}/\omega_{12} = 1/\gamma_0$ . Thus, observing  $\tilde{\beta} < \omega_{11}/\omega_{12}$  and  $\omega_{12} < 0$ , or  $\tilde{\beta}\omega_{12}/\omega_{11} > 1$ , violates the identifying restriction that  $\lambda \geq 0$  for only with a  $\lambda < 0$  can we get  $g(\beta) > 1$  when  $\beta < 0$  and hence  $\tilde{\beta} < \beta < 0$ .

The above arguments are illustrated in Figure 4.

### A.1.1 Bounds on $\lambda$

The bounds on  $\lambda$  are obtained by finding all the values of  $\lambda$  for which equation (40) has a solution for  $\beta$ . This equation implies

$$\tilde{\beta}(\omega_{11} - \omega_{12}\beta - \lambda\beta(\omega_{12} - \omega_{22}\beta)) - (1 - \lambda)(\omega_{11} - \omega_{12}\beta)\beta = 0,$$

or

$$\beta^2 \left( (1 - \lambda) \omega_{12} + \tilde{\beta} \lambda \omega_{22} \right) - \beta \left( (1 - \lambda) \omega_{11} + \tilde{\beta} (1 + \lambda) \omega_{12} \right) + \tilde{\beta} \omega_{11} = 0,$$

whose discriminant is

$$\begin{aligned} D(\lambda) &= \left( (1 - \lambda) \omega_{11} + \tilde{\beta} (1 + \lambda) \omega_{12} \right)^2 - 4 \tilde{\beta} \omega_{11} \left( (1 - \lambda) \omega_{12} + \tilde{\beta} \lambda \omega_{22} \right) \\ &= \lambda^2 \left( \omega_{11} - \tilde{\beta} \omega_{12} \right)^2 - 2 \lambda \left( \omega_{11}^2 - \tilde{\beta}^2 \omega_{12}^2 - 2 \tilde{\beta} \omega_{11} \omega_{12} + 2 \tilde{\beta}^2 \omega_{11} \omega_{22} \right) + \left( \omega_{11} - \tilde{\beta} \omega_{12} \right)^2. \end{aligned}$$

Hence, the identified set for  $\lambda$  corresponds to  $S_\lambda = \{\lambda : D(\lambda) \geq 0\}$ . This set is non-empty because  $D(0) \geq 0$ . It can be computed analytically and can take the following three shapes: (i)  $S_\lambda = \mathfrak{R}$  if  $D(\lambda) \geq 0$  for all  $\lambda \in \mathfrak{R}$ ; (ii)  $S_\lambda = (-\infty, \lambda_{lo}] \cup [\lambda_{up}, \infty)$  if  $\omega_{11} - \tilde{\beta} \omega_{12} \neq 0$ , where  $\lambda_{lo} < \lambda_{up}$  are the roots of  $D(\lambda) = 0$ ; and (iii)  $S_\lambda = (-\infty, \lambda_{lo}]$  if  $\omega_{11} - \tilde{\beta} \omega_{12} = 0$ , because  $\omega_{11}^2 - \left(\frac{\omega_{11}}{\omega_{12}}\right)^2 \omega_{12}^2 - 2 \left(\frac{\omega_{11}}{\omega_{12}}\right) \omega_{11} \omega_{12} + 2 \left(\frac{\omega_{11}}{\omega_{12}}\right)^2 \omega_{11} \omega_{22} = 2 \omega_{11}^2 \frac{\omega_{11} \omega_{22} - \omega_{12}^2}{\omega_{12}^2} > 0$ . If we also impose the restriction  $\lambda \in [0, 1]$ , then the identified set is  $S_\lambda \cap [0, 1]$ .

## A.2 Proof of Proposition 1

Define  $\bar{A}_{i2} := A_{i2}^* + A_{i2}$ ,  $i = 1, 2$  as the right blocks of  $\bar{A}$  that was defined in (19). Applying (Gourieroux et al., 1980, Theorem 1), coherency holds if and only if  $\det \bar{A}$  and  $\det A^*$  have the same sign. Without loss of generality, we can assume that  $A_{11}$  is nonsingular (this can always be achieved by reordering the variables in  $Y_t$ ). From (19), we have  $\det A^* = \det A_{11} \det (A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*)$  and  $\det \bar{A} = \det A_{11} \det (\bar{A}_{22} - A_{21} A_{11}^{-1} \bar{A}_{12})$  (Lütkepohl, 1996, p. 50 (6)). The coherency condition can be written as  $\det \bar{A} / \det A^* > 0$ , which, given that  $(\bar{A}_{22} - A_{21} A_{11}^{-1} \bar{A}_{12})$  and  $(A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*)$  are scalars, yields (20).<sup>21</sup>

## A.3 Proof of Proposition 2

Define  $\bar{A}_{i2} := A_{i2}^* + A_{i2}$ ,  $i = 1, 2$  as the right blocks of  $\bar{A}$  that was defined in (19). Also let  $Y_t^* := (Y'_{1t}, Y'_{2t})'$ . When the coherency condition (20) holds, the solution of (18) exists and is unique. It can be expressed as

$$Y_t^* = \begin{cases} CX_t + C^* X_t^* + u_t, & \text{if } D_t = 0 \\ \tilde{C} X_t + \tilde{C}^* X_t^* + \tilde{c} b + \tilde{u}_t, & \text{if } D_t = 1 \end{cases} \quad (41)$$

where

$$C = \bar{A}^{-1} B, \quad C^* = \bar{A}^{-1} B^*, \quad u_t = \bar{A}^{-1} \varepsilon_t \quad (42)$$

and

$$\tilde{C} = A^{*-1} B, \quad \tilde{C}^* = A^{*-1} B^*, \quad \tilde{c} = -A^{*-1} \begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix} b, \quad \tilde{u}_t = A^{*-1} \varepsilon_t. \quad (43)$$

---

<sup>21</sup>An alternative proof that does not rely on (Gourieroux et al., 1980, Theorem 1) can be found in the Supplementary Appendix.

Using the partitioned inverse formula, we obtain

$$\begin{aligned}\tilde{C}_1 &= (A_{11} - A_{12}^* A_{22}^{*-1} A_{21})^{-1} (B_1 - A_{12}^* A_{22}^{*-1} B_2) \\ \tilde{C}_2 &= (A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*)^{-1} (B_2 - A_{21} A_{11}^{-1} B_1)\end{aligned}$$

and

$$\begin{aligned}C_1 &= (A_{11} - \bar{A}_{12} \bar{A}_{22}^{-1} A_{21})^{-1} (B_1 - \bar{A}_{12} \bar{A}_{22}^{-1} B_2) \\ C_2 &= (\bar{A}_{22} - A_{21} A_{11}^{-1} \bar{A}_{12})^{-1} (B_2 - A_{21} A_{11}^{-1} B_1).\end{aligned}$$

Solving the latter for  $B_1$  and  $B_2$  yields

$$B_1 = A_{11} C_1 + \bar{A}_{12} C_2, \text{ and } B_2 = \bar{A}_{22} C_2 + A_{21} C_1.$$

Thus,

$$\begin{aligned}\tilde{C}_1 &= C_1 + (A_{11} - A_{12}^* A_{22}^{*-1} A_{21})^{-1} (\bar{A}_{12} - A_{12}^* A_{22}^{*-1} \bar{A}_{22}) C_2 = C_1 - \tilde{\beta} C_2, \text{ and} \\ \tilde{C}_2 &= (A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*)^{-1} (A_{22}^* - A_{21} A_{11}^{-1} A_{12}^* + A_{22} - A_{21} A_{11}^{-1} A_{12}) C_2 = \kappa C_2,\end{aligned}$$

where  $\kappa$  is given in (20). The exact same derivations apply to  $\tilde{C}^*$ , i.e.,

$$\tilde{C}_1^* = C_1^* - \tilde{\beta} C_2^*, \text{ and } \tilde{C}_2^* = \kappa C_2^*.$$

Next,

$$\begin{aligned}\tilde{c}_1 &= (A_{11} - A_{12}^* A_{22}^{*-1} A_{21})^{-1} (A_{12}^* A_{22}^{*-1} A_{22} - A_{12}) b = \tilde{\beta} b, \text{ and} \\ \tilde{c}_2 &= -\frac{A_{22} - A_{21} A_{11}^{-1} A_{12}}{A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*} b = (1 - \kappa) b.\end{aligned}$$

Finally,  $\tilde{u}_t = A^{*-1} \bar{A} u_t = (\tilde{u}_{1t}', \tilde{u}_{2t}')'$ , where

$$\tilde{u}_{1t} = u_{1t} - \tilde{\beta} u_{2t}, \text{ and } \tilde{u}_{2t} = \kappa u_{2t}.$$

Substituting back into (41), the reduced-form model for  $Y_{1t}$  becomes

$$\begin{aligned}Y_{1t} &= (1 - D_t) (C_1 X_t + C_1^* X_t^* + u_{1t}) \\ &\quad + D_t \left( (C_1 - \tilde{\beta} C_2) X_t + (C_1^* - \tilde{\beta} C_2^*) X_t^* + u_{1t} - \tilde{\beta} u_{2t} \right),\end{aligned}\tag{44}$$

and for  $Y_{2t}^*$  it is

$$Y_{2t}^* = C_2 X_t + C_2^* X_t^* + u_{2t} - (1 - \kappa) D_t (C_2 X_t + C_2^* X_t^* + u_{2t} - b). \quad (45)$$

Next, define

$$\tilde{Y}_{2t}^* := C_2 X_t + C_2^* X_t^* + u_{2t}, \quad (46)$$

and rewrite (45) as

$$\begin{aligned} Y_{2t}^* &= \tilde{Y}_{2t}^* - (1 - \kappa) D_t (\tilde{Y}_{2t}^* - b) \\ &= (1 - D_t) \tilde{Y}_{2t}^* + D_t (\kappa \tilde{Y}_{2t}^* + (1 - \kappa) b). \end{aligned} \quad (47)$$

Let  $q = \dim X_t$  denote the number of elements of  $X_t$  and define, for each  $i = 1, 2$ ,

$$\overline{C}_{ij} = \begin{cases} C_{ij}, & j \in \{1, q\} : X_{tj} \neq Y_{2,t-s} \text{ for all } s \in \{1, p\} \\ C_{ij} + C_{is}^*, & j \in \{1, q\} : X_{tj} = Y_{2,t-s}, \text{ for some } s \in \{1, p\}. \end{cases} \quad (48)$$

In other words,  $\overline{C}$  contains the original coefficients on all the regressors other than the lags of  $Y_{2t}$ , while the coefficients on the lags of  $Y_{2t}$  are augmented by the corresponding coefficients of the lags of  $Y_{2t}^*$ . For example, if  $p = 1$  and there are no other exogenous regressors  $X_{0t}$ , then, for  $i = 1, 2$ ,

$$C_i X_t + C_i^* X_t^* = C_{i1} Y_{1t-1} + C_{i2} Y_{2t-1} + C_i^* Y_{2t-1}^*,$$

so  $\overline{C}_i = (C_{i1}, C_{i2} + C_i^*)$ . Using (48), we can rewrite (46) as

$$\tilde{Y}_{2t}^* = \overline{C}_2 X_t + C_2^* \min(X_t^* - b, 0) + u_{2t}. \quad (49)$$

Now, observe that

$$\min(Y_{2t}^* - b, 0) = D_t (Y_{2t}^* - b) = \kappa D_t (\tilde{Y}_{2t}^* - b) = \kappa \min(\tilde{Y}_{2t}^* - b, 0)$$

So, letting  $\tilde{X}_t^*$  denote the lags of  $\tilde{Y}_{2t}^*$ , we have  $\min(X_t^* - b, 0) = \kappa \min(\tilde{X}_t^* - b, 0)$ , and consequently,

$$C^* \min(X_t^* - b, 0) = \overline{C}^* \min(\tilde{X}_t^* - b, 0),$$

where  $\overline{C}^* = \kappa C^*$ . Now, from (49) we have

$$\tilde{Y}_{2t}^* = \overline{C}_2 X_t + \overline{C}_2^* \min(\tilde{X}_t^* - b, 0) + u_{2t}.$$

Recall the definition of  $\bar{Y}_{2t}^*$  in (24):

$$\bar{Y}_{2t}^* := \bar{C}_2 X_t + \bar{C}_2^* \bar{X}_t^* + u_{2t},$$

where  $\bar{X}_t^* := (\bar{x}_{t-1}, \dots, \bar{x}_{t-p})'$ , and  $\bar{x}_t := \min(\bar{Y}_{2t}^* - b, 0)$ , with initial conditions  $\bar{x}_{-s} = \kappa^{-1} \min(Y_{2,-s}^* - b, 0)$ ,  $s = 0, \dots, p-1$ . It follows that  $\min(\bar{X}_t^* - b, 0) = \bar{X}_t^*$  for all  $t \geq 1$ , so that  $\tilde{Y}_{2t}^* = \bar{Y}_{2t}^*$ . Substituting  $\bar{Y}_2^*$  for  $\tilde{Y}_2^*$  in (47), we get (25). Using the reparametrization (48) and the relationship between  $X_t^*$  and  $\bar{X}_t^*$  in (44), we obtain (22).

Finally, from eq. (45), it follows that the event  $Y_{2t}^* < b$  is equivalent to

$$b + \kappa (C_2 X_t + C_2^* X_t^* + u_{2t} - b) < b,$$

which, since  $\kappa > 0$  by the coherency condition (20), is equivalent to

$$u_{2t} < b - C_2 X_t - C_2^* X_t^*. \quad (50)$$

Using the definition (24), and (48), the inequality (50) can be written as  $\bar{Y}_{2t}^* < b$ , which establishes (23).

**Comment:** Note that  $\kappa$  appears in the reduced form only multiplicatively with  $C^*$ , so  $\kappa$  and  $C^*$  are not separately identified, only  $\bar{C}^* = \kappa C^*$  is. The reparametrization from  $C$  to  $\bar{C}$  is convenient because  $\bar{C}$  is identified independently of  $\kappa$ , while  $C, C^*$  and  $\kappa$  are not separately identified.

#### A.4 Proof of Proposition 3

We solve  $u_t = \bar{A}^{-1} \varepsilon_t$  using the partitioned inverse formula to get

$$u_{1t} = \left( A_{11} - \bar{A}_{12} \bar{A}_{22}^{-1} A_{21} \right)^{-1} \left( \varepsilon_{1t} - \bar{A}_{12} \bar{A}_{22}^{-1} \varepsilon_{2t} \right) \quad (51)$$

$$u_{2t} = \left( \bar{A}_{22} - A_{21} A_{11}^{-1} \bar{A}_{12} \right)^{-1} \left( \varepsilon_{2t} - A_{21} A_{11}^{-1} \varepsilon_{1t} \right). \quad (52)$$

Using the definitions

$$\begin{aligned} \bar{\beta} &:= -A_{11}^{-1} \bar{A}_{12}, & \bar{\gamma} &:= -\bar{A}_{22}^{-1} A_{21}, \\ \bar{\varepsilon}_{1t} &:= A_{11}^{-1} \varepsilon_{1t}, & \bar{\varepsilon}_{2t} &:= \bar{A}_{22}^{-1} \varepsilon_{2t}, \end{aligned}$$

we can rewrite (51)-(52) as (30)-(31).

Note that

$$\begin{aligned}\bar{\varepsilon}_{1t} &= A_{11}^{-1} (A_{11}u_{1t} + \bar{A}_{12}u_{2t}) = u_{1t} - \bar{\beta}u_{2t}, \\ \bar{\varepsilon}_{2t} &= \bar{A}_{22}^{-1} (A_{21}u_{1t} + \bar{A}_{22}u_{2t}) = -\bar{\gamma}u_{1t} + u_{2t},\end{aligned}$$

so,

$$\begin{aligned}\text{var}(\bar{\varepsilon}_{1t}) &= (I_{k-1}, -\bar{\beta}) \Omega (I_{k-1}, -\bar{\beta})', \\ \text{var}(\bar{\varepsilon}_{2t}) &= (-\bar{\gamma}, 1) \Omega (-\bar{\gamma}, 1)',\end{aligned}$$

and

$$\begin{aligned}\text{cov}(\bar{\varepsilon}_{1t}, \bar{\varepsilon}_{2t}) &= (I_{k-1}, -\bar{\beta}) \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega'_{12} & \Omega_{22} \end{pmatrix} (-\bar{\gamma}, 1)' \\ &= -(\Omega_{11} - \bar{\beta}\Omega'_{12}) \bar{\gamma}' + \Omega_{12} - \bar{\beta}\Omega_{22} = 0.\end{aligned}$$

The last equation identifies  $\bar{\gamma}$  given  $\bar{\beta}$ . Specifically,

$$\bar{\gamma}' = (\Omega_{11} - \bar{\beta}\Omega'_{12})^{-1} (\Omega_{12} - \bar{\beta}\Omega_{22}).$$

## B Computational details

### B.1 Likelihood

To compute the likelihood, we need to obtain the prediction error densities. The first step is to write the model in state-space form. Define

$$s_t = \begin{pmatrix} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \end{pmatrix}, \quad \mathbf{y}_t = \begin{pmatrix} Y_t \\ \bar{Y}_{2t}^* \end{pmatrix},$$

and write the state transition equation as

$$s_t = F(s_{t-1}, u_t; \psi) = \begin{pmatrix} F_1(s_{t-1}, u_t; \psi) \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{pmatrix}, \quad (53)$$

$$F_1(s_{t-1}, u_t; \psi) = \begin{pmatrix} \bar{C}_1 X_t + \bar{C}_1^* \bar{X}_t^* + u_{1t} - \tilde{\beta} D_t (\bar{C}_2 X_t + \bar{C}_2^* \bar{X}_t^* + u_{2t} - b) \\ \max \left( b, \bar{C}_2 X_t + \bar{C}_2^* \bar{X}_t^* + u_{2t} \right) \\ \bar{C}_2 X_t + \bar{C}_2^* \bar{X}_t^* + u_{2t} \end{pmatrix},$$

and the observation equation as

$$Y_t = \begin{pmatrix} I_k & 0_{k \times 1 + (p-1)(k+1)} \end{pmatrix} s_t. \quad (54)$$

Next, I will derive the predictive density and mass functions. With Gaussian errors, the joint predictive density of  $Y_t$  corresponding to the observations with  $D_t = 0$  can be written as:

$$f_0(Y_t | s_{t-1}, \psi) := |\Omega|^{-1/2} \exp \left\{ -\frac{1}{2} \text{tr} \left( \left( Y_t - \bar{C} X_t - \bar{C}^* \bar{X}_t^* \right) \left( Y_t - \bar{C} X_t - \bar{C}^* \bar{X}_t^* \right)' \Omega^{-1} \right) \right\}. \quad (55)$$

At  $D_t = 1$ , the predictive density of  $Y_{1t}$  can be written as:

$$f_1(Y_{1t} | s_{t-1}, \psi) := |\Xi_1|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (Y_{1t} - \mu_{1t})' \Xi_1^{-1} (Y_{1t} - \mu_{1t}) \right] \quad (56)$$

$$\mu_{1t} := \tilde{\beta} b + (\bar{C}_1 - \tilde{\beta} \bar{C}_2) X_t + (\bar{C}_1^* - \tilde{\beta} \bar{C}_2^*) \bar{X}_t^* \quad (57)$$

$$\Xi_1 := \Omega_{1.2} + \tilde{\delta} \tilde{\delta}' \tau^2 = \begin{pmatrix} I_{k-1} & -\tilde{\beta} \end{pmatrix} \Omega \begin{pmatrix} I_{k-1} \\ -\tilde{\beta}' \end{pmatrix}, \quad \tilde{\delta} = \Omega_{12} \omega_{22}^{-1} - \tilde{\beta}, \quad (58)$$

$\Omega_{1.2} = \Omega_{11} - \Omega_{12} \omega_{22}^{-1} \Omega_{21}$ , and  $\tau = \sqrt{\omega_{22}}$ . Next,

$$u_{2t} | Y_{1t}, s_{t-1} \sim N(\mu_{2t}, \tau_2^2), \text{ with} \quad (59)$$

$$\mu_{2t} := \tau^2 \tilde{\delta}' \Xi_1^{-1} (Y_{1t} - \mu_{1t}), \quad \tau_2 = \tau \sqrt{\left( 1 - \tau^2 \tilde{\delta}' \Xi_1^{-1} \tilde{\delta} \right)}. \quad (60)$$

Hence,

$$\Pr(D_t = 1 | Y_{1t}, s_{t-1}, \psi) = \Phi \left( \frac{b - \bar{C}_2 X_t - \bar{C}_2^* \bar{X}_t^* - \mu_{2t}}{\tau_2} \right). \quad (61)$$

In the case of the KSVAR model, there are no latent lags ( $\bar{C}^* = 0$ ,  $\bar{C} = C$ ), so the log-likelihood is available analytically:

$$\log L(\psi) = \sum_{t=1}^T \left[ (1 - D_t) \log f_0(Y_t | s_{t-1}, \psi) + D_t \log \left( f_1(Y_{1t} | s_{t-1}, \psi) \Phi \left( \frac{b - C_2 X_t - \mu_{2t}}{\tau_2} \right) \right) \right] \quad (62)$$

where  $f_0(Y_t | s_{t-1}, \theta)$  and  $f_1(Y_{1t} | s_{t-1}, \theta)$  are given by (55) and (56), resp., with  $\bar{C}^* = 0$ .

The likelihood for the unrestricted CKSVAR ( $\bar{C}^* \neq 0$ ) is not available analytically, but it can be computed approximately by simulation (particle filtering). I provide two different simulation algorithms. The first is a sequential importance sampler (SIS), proposed originally by Lee (1999) for the univariate dynamic Tobit model. It is extended here to the CKSVAR model. The second algorithm is

a fully adapted particle filter (FAPF), which is a sequential importance resampling algorithm designed to address the sample degeneracy problem. It is proposed by Malik and Pitt (2011) and is a special case of the auxiliary particle filter developed by Pitt and Shephard (1999).

Both algorithms require sampling from the predictive density of  $\bar{Y}_{2t}^*$  conditional on  $Y_{1t}, D_t = 1$  and  $s_{t-1}$ . From (24) and (59), we see that this is a truncated Normal with original mean  $\mu_{2t}^* = \bar{C}_2 X_t + \bar{C}_2^* \bar{X}_t^* + \mu_{2t}$  and standard deviation  $\tau_2$ , where  $\mu_{2t}, \tau_2$  are given in (60), i.e.,

$$f_2(Y_{2t}^* | Y_{1t}, D_t = 1, s_{t-1}, \psi) = TN\left(\mu_{2t}^*, \tau_2, \bar{Y}_{2t}^* < b\right) \quad (63)$$

Draws from this truncated distribution can be obtained using, for instance, the procedure in Lee (1999). Let  $\xi_t^{(j)} \sim U[0, 1]$  be *i.i.d.* uniform random draws,  $j = 1, \dots, M$ . Then, a draw from  $\bar{Y}_{2t}^* | Y_{1t}, s_{t-1}, \bar{Y}_{2t}^* < b$  is given by

$$\bar{Y}_{2t}^{*(j)} = \mu_{2t}^* + \tau_2 \Phi^{-1} \left[ \xi_t^{(j)} \Phi \left( \frac{b - \mu_{2t}^*}{\tau_2} \right) \right]. \quad (64)$$

**Algorithm 1 (SIS)** *Sequential Importance Sampler*

1. Initialization. For  $j = 1 : M$ , set  $W_0^j = 1$  and  $s_0^j = (\mathbf{y}_0^j, \dots, \mathbf{y}_{-p+1}^j)$ , with  $\mathbf{y}_{-s}^j = (Y_{0'}', Y_{2,0})'$ , for  $s = 0, \dots, p-1$ . (in other words, initialize  $\bar{Y}_{2,-s}^*$  at the observed values of  $Y_{2,-s}$ ).

2. Recursion. For  $t = 1 : T$ :

(a) For  $j = 1 : M$ , compute the incremental weights

$$w_{t-1|t}^j = p(Y_t | s_{t-1}^j, \psi) = \begin{cases} f_0(Y_t | s_{t-1}^j, \psi), & \text{if } D_t = 0 \\ f_1(Y_{1t} | s_{t-1}^j, \psi) \Pr(D_t = 1 | Y_{1t}, s_{t-1}^j, \psi), & \text{if } D_t = 1 \end{cases}$$

where  $f_0, f_1$ , and  $\Pr(D_t = 1 | Y_{1t}, s_{t-1}; \psi)$  are given by (55), (56), and (61), resp., and

$$S_t = \frac{1}{M} \sum_{j=1}^M w_{t-1|t}^j W_{t-1}^j$$

(b) Sample  $s_t^j$  randomly from  $p(s_t | s_{t-1}^j, Y_t)$ . That is,  $s_t^j = (\mathbf{y}_t^j, \mathbf{y}_{t-1}^j, \dots, \mathbf{y}_{t-p}^j)$  where  $\mathbf{y}_t^j = (Y_t', \bar{Y}_{2t}^{*(j)})$  and  $\bar{Y}_{2t}^{*(j)}$  is a draw from  $f_2(Y_{2t}^* | Y_{1t}, D_t = 1, s_{t-1}^j, \psi)$  using (64).

(c) Update the weights:

$$W_t^j = \frac{w_{t-1|t}^j W_{t-1}^j}{S_t}.$$

3. Likelihood approximation

$$\log \hat{p}(Y_T | \psi) = \sum_{t=1}^T \log S_t$$



If the draws  $\xi_t^{(j)}$  are kept fixed across different values of  $\psi$ , the simulated likelihood in step 3 is smooth. Note that when  $k = 1$  and  $Y_t = Y_{2t}$  (no  $Y_{1t}$  variables), the model reduces to a univariate dynamic Tobit model, and Algorithm 1 reduces exactly to the sequential importance sampler proposed by Lee (1999). A possible weakness of this algorithm is sample degeneracy, which arises when all but a few weights  $W_t^j$  are zero. To gauge possible sample degeneracy, we can look at the effective sample size (ESS), as recommended by Herbst and Schorfheide (2015)

$$ESS_t = \frac{M}{\frac{1}{M} \sum_{j=1}^M (W_t^j)^2}. \quad (65)$$

Next, I turn to the FAPF algorithm.

**Algorithm 2 (FAPF)** *Fully Adapted Particle Filter*

1. Initialization. For  $j = 1 : M$ , set  $s_0^j = (\mathbf{y}_0^j, \dots, \mathbf{y}_{-p+1}^j)$ , with  $\mathbf{y}_{-s}^j = (Y_0', Y_{2,0})'$ , for  $s = 0, \dots, p-1$ . (in other words, initialize  $\bar{Y}_{2,-s}^*$  at the observed values of  $Y_{2,-s}$ ).

2. Recursion. For  $t = 1 : T$ :

(a) For  $j = 1 : M$ , compute

$$w_{t-1|t}^j = p(Y_t | s_{t-1}^j, \psi) = \begin{cases} f_0(Y_t | s_{t-1}^j, \psi), & \text{if } D_t = 0 \\ f_1(Y_{1t} | s_{t-1}^j, \psi) \Pr(D_t = 1 | Y_{1t}, s_{t-1}^j, \psi), & \text{if } D_t = 1 \end{cases}$$

where  $f_0, f_1$ , and  $\Pr(D_t = 1 | Y_{1t}, s_{t-1}; \psi)$  are given by (55), (56), and (61), resp., and

$$\pi_{t-1|t}^j = \frac{w_{t-1|t}^j}{\sum_{j=1}^M w_{t-1|t}^j}.$$

(b) For  $j = 1 : M$ , sample  $k_j$  randomly from the multinomial distribution  $\{j, \pi_{t-1|t}^j\}$ . Then, set  $\tilde{s}_{t-1}^j = s_{t-1}^{k_j}$  (this applies only to the elements in  $s_{t-1}^j$  that correspond to  $X_t^{*j}$ , since all the other elements are observed and constant across all  $j$ . That is,  $\tilde{s}_{t-1}^j = (\tilde{\mathbf{y}}_{t-1}^j, \dots, \tilde{\mathbf{y}}_{t-p}^j)$ ,  $\tilde{\mathbf{y}}_{t-s}^j = (Y_{t-1}', \bar{Y}_{2,t-s}^{*(k_j)})$ ,  $s = 1, \dots, p$ .)

(c) For  $j = 1 : M$ , sample  $s_t^j$  randomly from  $p(s_t | \tilde{s}_{t-1}^j, Y_t)$ . That is,  $s_t^j = (\mathbf{y}_t^j, \tilde{\mathbf{y}}_{t-1}^j, \dots, \tilde{\mathbf{y}}_{t-p}^j)$  where  $\mathbf{y}_t^j = (Y_t', \bar{Y}_{2t}^{*(j)})$  and  $\bar{Y}_{2t}^{*(j)}$  is a draw from  $f_2(Y_{2t}^* | Y_{1t}, D_t = 1, \tilde{s}_{t-1}^j, \psi)$  using (64).

3. Likelihood approximation

$$\ln \hat{p}(Y_T | \psi) = \sum_{t=1}^T \ln \left( \frac{1}{M} \sum_{j=1}^M w_{t-1|t}^j \right)$$

Many of the generic particle filtering algorithms used in the macro literature, described in Herbst and Schorfheide (2015), are inapplicable in a censoring context because of the absence of measurement error in the observation equation. It is, of course, possible to introduce a small measurement error in  $Y_{2t}$ , so that the constraint  $Y_{2t} \geq b$  is not fully respected, but there is no reason to expect other particle filters discussed in Herbst and Schorfheide (2015) to estimate the likelihood more accurately than the FAPF algorithm described above.

Moments or quantiles of the filtering or smoothing distribution of any function  $h(\cdot)$  of the latent states  $s_t$  can be computed using the drawn sample of particles. When we use Algorithm 2, simple average or quantiles of  $h(s_t^j)$  produce the requisite average or quantiles of  $h(s_t)$  conditional on  $Y_1, \dots, Y_t$  (the filtering density). For particles generated using Algorithm 1, we need to take weighted averages using the importance sampling weights  $W_t$ . Smoothing estimates of  $h(s_t^j)$  can be obtained using weights  $W_T$ .

## B.2 Computation of the identified set

Substitute for  $\bar{\gamma}$  in (37) using Proposition 3 to get

$$\tilde{\beta} = (1 - \xi) \left( I - \xi \bar{\beta} \left( \Omega'_{12} - \Omega_{22} \bar{\beta}' \right) \left( \Omega_{11} - \Omega_{12} \bar{\beta}' \right)^{-1} \right)^{-1} \bar{\beta}. \quad (66)$$

For each value of  $\xi \in [0, 1)$ , the above equation defines a correspondence from  $\Re^{k-1}$  to  $\Re^{k-1}$ . The range of  $\bar{\beta}$  can then be obtained numerically by solving (66) for  $\bar{\beta}$  as a function of the reduced-form parameters and  $\xi$  for each value of  $\xi$ , and gathering all the solutions in the set. It is shown in the Supplementary Appendix that (66) can be written as

$$\tilde{\beta} - \bar{A} \bar{\beta} + \bar{\beta} \bar{\beta}' \tilde{b} = 0, \quad (67)$$

where

$$\begin{aligned} \tilde{b} &= \Omega_{11}^{-1} \left( (\Omega_{22} - \Omega'_{12} \Omega_{11}^{-1} \Omega_{12}) I_{k-1} + \Omega_{12} \Omega'_{12} \Omega_{11}^{-1} \right) \tilde{\beta} \xi, \text{ and} \\ \bar{A} &= \tilde{\beta} \Omega'_{12} \Omega_{11}^{-1} + \left( \xi \Omega'_{12} \tilde{\beta} + 1 - \xi \right) I_{k-1}. \end{aligned}$$

Defining  $z := \tilde{b}' \bar{\beta}$ , it is further shown in the Supplementary Appendix that premultiplying equation (67) by  $\tilde{b}'$  yields the equation

$$\begin{aligned} 0 &= \tilde{b}' \tilde{\beta} \det(C_0(z)) + \tilde{b}' \bar{A} \tilde{b} \left( \tilde{b}' \tilde{b} \right)^{-1} z \det(C_0(z)) \\ &\quad - \tilde{b}' \bar{A} \tilde{b}_{\perp} C_0(z)^{adj} c_1(z) + \det(C_0(z)) z^2, \end{aligned} \quad (68)$$

where

$$\begin{aligned} C_0(z) &:= -\left(\tilde{b}'_{\perp} \tilde{A} \tilde{b}_{\perp} + \tilde{b}'_{\perp} \tilde{b}_{\perp} z\right), \\ c_1(z) &:= \tilde{b}'_{\perp} \tilde{\beta} + \tilde{b}'_{\perp} \tilde{A} \tilde{b} \left(\tilde{b}' \tilde{b}\right)^{-1} z, \end{aligned}$$

$\tilde{b}_{\perp}$  is a  $k-1 \times k-2$  matrix such that  $\tilde{b}'_{\perp} \tilde{b} = 0$ , and  $\det(C)$  and  $C^{adj}$  denote the determinant and the adjoint of a square matrix  $C$ , respectively. This is a polynomial equation of order  $k$  and has at most  $k$  real roots, denoted  $z_i$ , say. Then, the solutions for  $\bar{\beta}$  are given by

$$\bar{\beta}_i = \tilde{b} \left(\tilde{b}' \tilde{b}\right)^{-1} z_i + \tilde{b}_{\perp} C_0(z_i)^{-1} c_1(z_i). \quad (69)$$

provided that  $\det(C_0(z)) \neq 0$ .

An algorithm for obtaining the identified set of the IRF (38) is as follows.

**Algorithm 3 (ID set)** *Discretize the space  $(0, 1)$  into  $R$  equidistant points.*

*For each  $r = 1 : R$ , set  $\xi_r = \frac{r}{R+1}$  and solve equation (68).*

*1. If no solution exists, proceed to the next  $r$ .*

*2. If  $0 < q_r \leq k$  solutions exist, denote them  $z_{i,r}$ , and, for each  $i = 1 : q_r$ ,*

$$\begin{aligned} (a) \text{ derive } \bar{\beta}_{i,r} \text{ from (69), } \bar{\gamma}_{i,r} &= \left(\Omega'_{12} - \Omega_{22} \bar{\beta}'_{i,r}\right) \left(\Omega_{11} - \Omega_{12} \bar{\beta}_{i,r}\right)^{-1}, \\ \bar{A}_{22,i,r}^{-1} &= \sqrt{(-\bar{\gamma}_{i,r}, 1) \Omega (-\bar{\gamma}_{i,r}, 1)'} \text{, and } \Xi_{1,i,r} = (I_{k-1}, -\bar{\beta}_{i,r}) \Omega (I_{k-1}, -\bar{\beta}_{i,r})'; \end{aligned}$$

*(b) for  $j = 1 : M$ ,*

*i. draw independently  $\bar{\varepsilon}_{1t,i,r}^j \sim N(0, \Xi_{1,i,r})$  and  $u_{t+h}^j \sim N(0, \Omega)$  for  $h = 1, \dots, H$ ;*

*ii. for any scalar  $\varsigma$ , set*

$$\begin{aligned} u_{1t,i,r}^j(\varsigma) &= (I_{k-1} - \bar{\beta}_{i,r} \bar{\gamma}_{i,r})^{-1} \left(\bar{\varepsilon}_{1t,i,r}^j - \bar{\beta}_{i,r} \varsigma\right) \\ u_{2t,i,r}^j(\varsigma) &= (1 - \bar{\gamma}_{i,r} \bar{\beta}_{i,r})^{-1} \left(\varsigma - \bar{\gamma}_{i,r} \bar{\varepsilon}_{1t,i,r}^j\right), \end{aligned}$$

*and compute  $Y_{t,i,r}^j(\varsigma)$  using (22)-(23) with  $u_{t,i,r}^j(\varsigma)$  in place of  $u_t$ , and iterate forward to obtain  $Y_{t+h,i,r}^j(\varsigma)$  using  $u_{t+h}^j$  computed in step i. Set  $\varsigma = 1$  for a one-unit (e.g., 100 basis points) impulse to the policy shock  $\bar{\varepsilon}_{2t}$ , or  $\varsigma = \bar{A}_{22,i,r}^{-1}$  for a one-standard deviation impulse.*

*(c) compute*

$$\widehat{IRF}_{h,t,i,r}(\varsigma) = \frac{1}{M} \sum_{j=1}^M \left(Y_{t+h,i,r}^j(\varsigma) - Y_{t+h,i,r}^j(0)\right).$$

*The identified set is given by the collection of  $\widehat{IRF}_{h,t,i,r}(\varsigma)$  over  $i = 1 : q_r$ ,  $r = 1 : R$ , and the single point-identified IRF at  $\xi = 0$ .*

Mnemonic	Description
$\tau$	st. dev. of reduced form error $u_{2t}$ in $Y_{2t}$ (constrained variable)
Eq. 3	reduced form equation for $Y_{2t}$
Eq. j	red. form equation for $Y_{1j,t}$ , $j = 1, 2$ (unconstrained variables)
$\tilde{\beta}_j$	coefficient on kink in eq. j
eq. i Y1j_1	coefficient of $Y_{1j,t-1}$ in eq. i
eq. i Y2_1	coefficient of $Y_{2,t-1}$ in eq. i
eq. i lY2_1	coefficient of $\min(Y_{2,t-1}^* - b, 0)$ in eq. i
$\delta_j$	coefficient of regression of $u_{1j,t}$ (red. form error in Eq. j) on $u_{2t}$
Ch_ij	(i,j) element of Choleski factor of $\Omega_{1,2}$

Table 2: Parameter notation in reported simulation results.

## C Numerical results

This section provides Monte-Carlo evidence on the finite-sample properties of the proposed estimators and tests. The data generating process (DGP) is a trivariate VAR(1), given by equations (16) and (17). I consider three different DGPs corresponding to the CKSVAR, KSVAR and CSVAR models, respectively. In all three DGPs, the following parameters are set to the same values: the contemporaneous coefficients are  $A_{11} = I_2$ ,  $A_{12} = A_{12}^* = 0_{2 \times 1}$ ,  $A_{22}^* = 1$  and  $A_{22} = 0$ ; the intercepts are set to zero,  $B_{10} = 0_{2 \times 1}$  and  $B_{20} = 0$ ; the coefficients on the lags are  $B_{1,1} = (\rho I_2, 0)$ ,  $B_{1,1}^* = 0_{2 \times 1}$ ,  $B_{2,1} = (0_{1 \times 2}, B_{22,1})$ , with  $\rho = 0.5$ . Finally, each of the three DGPs is determined as follows. DGP1:  $B_{22,1} = B_{2,1}^* = 0$  (both KSVAR and CSVAR restrictions hold, since lags of  $Y_{2,t}$  and  $Y_{2,t}^*$  all have zero coefficients); DGP2:  $B_{22,1} = \rho$ ,  $B_{2,1}^* = 0$  (KSVAR restrictions hold but CSVAR restrictions do not); DGP3:  $B_{22,1} = 0$ ,  $B_{2,1}^* = \rho$  (CSVAR restrictions hold but KSVAR restrictions do not). The setting of the autoregressive coefficient  $\rho = 0.5$  leads to a lower degree of persistence than is typically observed in macro data (e.g., in the Stock and Watson, 2001, application, the three largest roots are 0.97, 0.97 and 0.8), because I want to avoid confounding any possible finite-sample issues arising from the ZLB with well-known problems of bias and size distortion due to strong persistence (near unit roots) in the data. Finally, the bound on  $Y_{2t}$  is set to  $b = 0$ , the sample size is  $T = 250$ , the initial conditions are set to 0 and the number of Monte Carlo replications is 1000. In all cases, the CKSVAR and CSVAR likelihoods are computed using SIS with  $R = 1000$  particles. The notation for the reported parameters is given in Table 2.

Figures 5, 6 and 7 report the sampling distribution of the ML estimators of the reduced-form parameters in Proposition 2 for the CKSVAR, KSVAR and CSVAR models, respectively, under DGP1 (all three models are correctly specified). The sampling densities appear to be very close to the superimposed Normal approximations, indicating that the Normal asymptotic approximation is fairly accurate.

Tables 3, 4 and 5 report moments of the sampling distributions of the above mentioned estimators. We notice no discernible biases. Unreported results with  $T = 100$  and  $T = 1000$  indicate that the

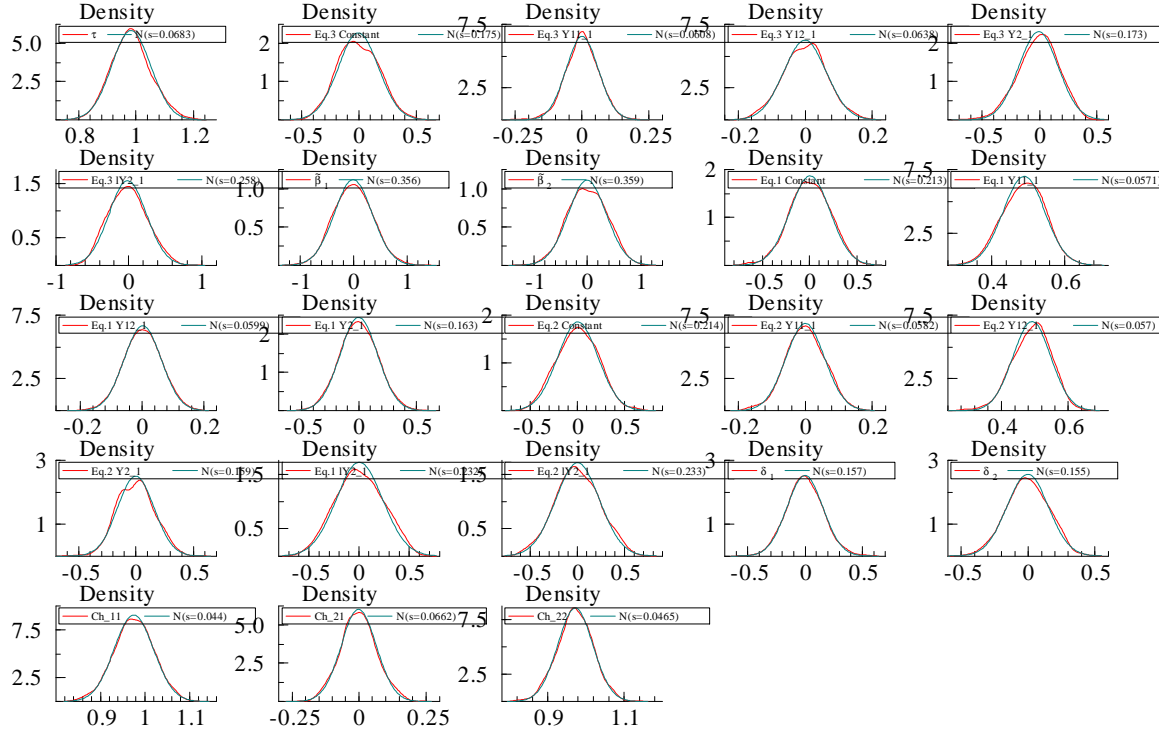


Figure 5: Sampling distribution of reduced-form coefficients of CKSVAR(1) in DGP1.  $T = 250, 1000$  Monte Carlo replications. Parameter names described in Table 2.

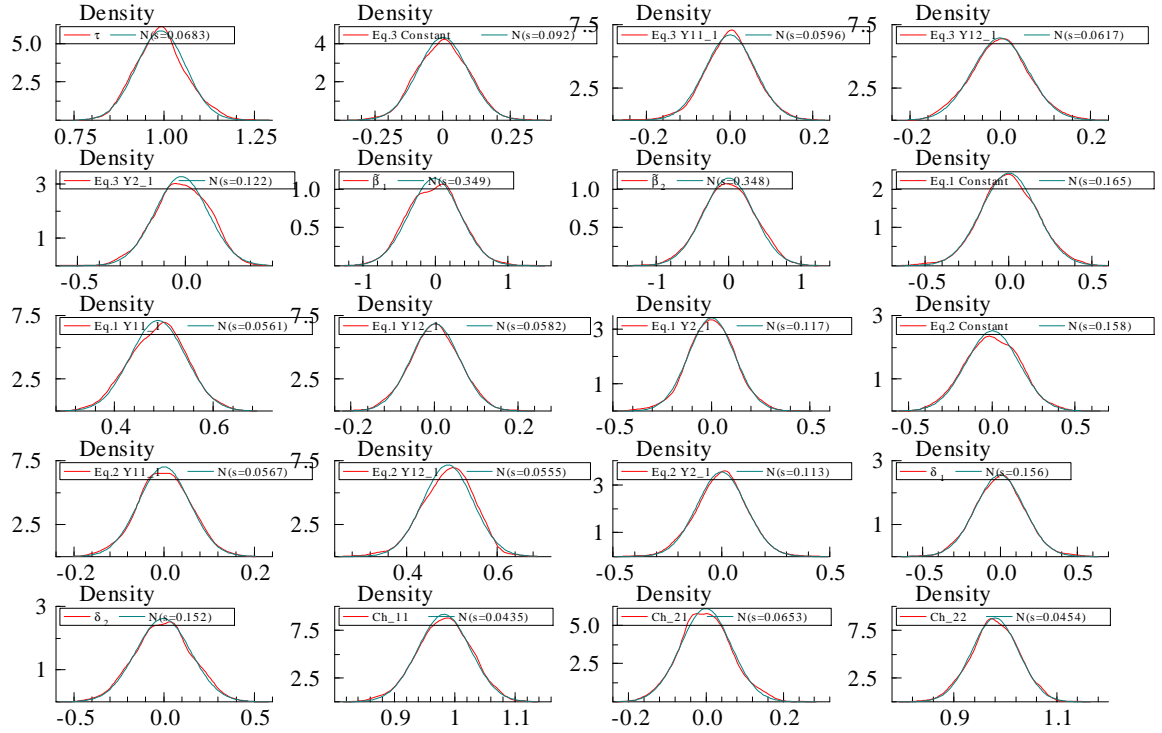


Figure 6: Sampling distribution of reduced-form coefficients of KSVAR(1) in DGP1.  $T = 250, 1000$  Monte Carlo replications. Parameter names described in Table 2.

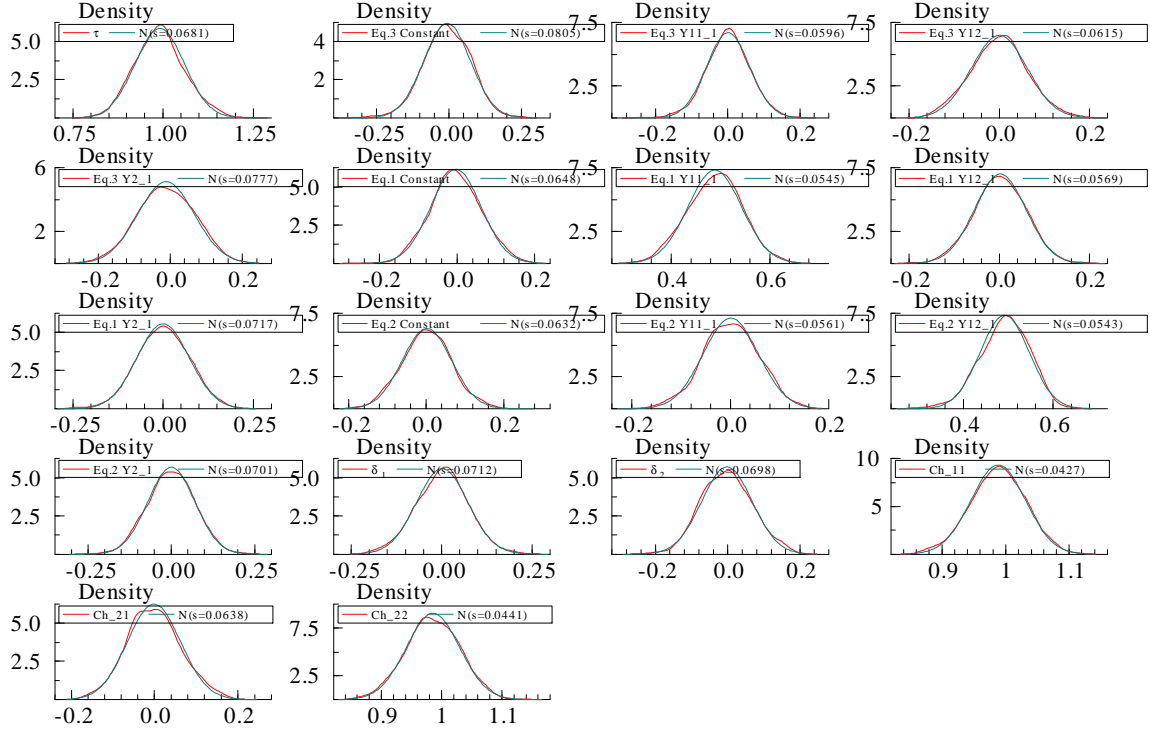


Figure 7: Sampling distribution of reduced-form coefficients of CSVAR(1) in DGP1.  $T = 250$ , 1000 Monte Carlo replications. Parameter names described in Table 2.

RMSE declines at rate  $\sqrt{T}$  in accordance with asymptotic theory, see the Supplementary Appendix. It is noteworthy that the estimators of  $\tilde{\beta}$  in CKSVAR and KSVAR have substantially larger RMSE than the estimators of the other parameters.

Next, I turn to the properties of the LR test of KSVAR against CKSVAR and CSVAR against CKSVAR. The former hypothesis involves three restrictions (exclusion of the latent lag  $Y_{2,t-1}^*$  from each of the three equations), so the LR statistic is asymptotically distributed as  $\chi_3^2$  under the null. The latter hypothesis involves five restrictions (exclusion of the observed lag  $Y_{2,t-1}$  from each of the three equations, plus  $\tilde{\beta} = 0$ ), and the LR statistic is asymptotically distributed as  $\chi_5^2$ . Table 6 reports the rejection frequencies of the LR tests for each of the two hypotheses in each of the three DGPs at three significance levels: 10%, 5% and 1%. In addition to the asymptotic tests, I also report the rejection frequency of the tests using parametric bootstrap critical values. The parametric bootstrap uses draws of Normal errors and the estimated reduced-form parameters to generate the bootstrap samples of  $Y_t$  and  $Y_{2,t}^*$ . The Monte Carlo rejection frequencies are computed using the “warp-speed” method method of Giacomini et al. (2013). Note that both null hypotheses hold under DGP1, but only the KSVAR is valid under DGP2 and only the CSVAR is valid under DGP3. For convenience, I indicate the rejection frequencies under the alternative in bold in the table.

There is evidence that the LR tests reject too often under  $H_0$  relative to their nominal level when we use asymptotic critical values. Moreover, the size distortions are very similar across null hypotheses and DGPs. Unreported results show that size distortion eventually disappears as the

ML-CKSVAR	true	mean	bias	sd	RMSE
$\tau$	1.000	0.983	-0.017	0.068	0.070
Eq.3 Constant	0.000	-0.006	-0.006	0.175	0.176
Eq.3 Y11_1	0.000	0.000	0.000	0.061	0.061
Eq.3 Y12_1	0.000	-0.000	-0.000	0.064	0.064
Eq.3 Y2_1	0.000	-0.010	-0.010	0.173	0.173
Eq.3 lY2_1	0.000	-0.013	-0.013	0.258	0.258
$\tilde{\beta}_1$	-0.000	-0.008	-0.008	0.356	0.356
$\tilde{\beta}_2$	0.000	-0.004	-0.004	0.359	0.359
Eq.1 Constant	0.000	0.002	0.002	0.213	0.213
Eq.1 Y11_1	0.500	0.489	-0.011	0.057	0.058
Eq.1 Y12_1	0.000	0.002	0.002	0.060	0.060
Eq.1 Y2_1	0.000	-0.002	-0.002	0.163	0.163
Eq.2 Constant	0.000	0.005	0.005	0.214	0.214
Eq.2 Y11_1	0.000	0.000	0.000	0.058	0.058
Eq.2 Y12_1	0.500	0.491	-0.009	0.057	0.058
Eq.2 Y2_1	0.000	-0.001	-0.001	0.159	0.159
Eq.1 lY2_1	0.000	0.005	0.005	0.232	0.232
Eq.2 lY2_1	0.000	0.002	0.002	0.233	0.233
$\delta_1$	0.000	-0.001	-0.001	0.157	0.157
$\delta_2$	0.000	-0.004	-0.004	0.155	0.155
Ch_11	1.000	0.975	-0.025	0.044	0.051
Ch_21	0.000	-0.001	-0.001	0.066	0.066
Ch_22	1.000	0.972	-0.028	0.046	0.054

Table 3: Moments of sampling distribution of ML estimator of the parameters CKSVAR(1), under DGP1.  $T = 250$ , 1000 MC replications. Parameter names described in Table 2.

ML-KSVAR	true	mean	bias	sd	RMSE
$\tau$	1.000	0.992	-0.008	0.068	0.069
Eq.3 Constant	0.000	0.001	0.001	0.092	0.092
Eq.3 Y11_1	0.000	0.001	0.001	0.060	0.060
Eq.3 Y12_1	0.000	-0.000	-0.000	0.062	0.062
Eq.3 Y2_1	0.000	-0.019	-0.019	0.122	0.124
$\tilde{\beta}_1$	-0.000	-0.013	-0.013	0.349	0.349
$\tilde{\beta}_2$	0.000	-0.001	-0.001	0.348	0.348
Eq.1 Constant	0.000	0.001	0.001	0.165	0.165
Eq.1 Y11_1	0.500	0.488	-0.012	0.056	0.057
Eq.1 Y12_1	0.000	0.002	0.002	0.058	0.058
Eq.1 Y2_1	0.000	-0.000	-0.000	0.117	0.117
Eq.2 Constant	0.000	0.003	0.003	0.158	0.158
Eq.2 Y11_1	0.000	0.001	0.001	0.057	0.057
Eq.2 Y12_1	0.500	0.492	-0.008	0.055	0.056
Eq.2 Y2_1	0.000	-0.000	-0.000	0.113	0.113
$\delta_1$	0.000	-0.003	-0.003	0.156	0.156
$\delta_2$	0.000	-0.003	-0.003	0.152	0.152
Ch_11	1.000	0.982	-0.018	0.044	0.047
Ch_21	0.000	-0.000	-0.000	0.065	0.065
Ch_22	1.000	0.980	-0.020	0.045	0.050

Table 4: Moments of sampling distribution of ML estimator of the parameters KSVAR(1), under DGP1.  $T = 250$ , 1000 MC replications. Parameter names described in Table 2.

ML-CSVAR	true	mean	bias	sd	RMSE
$\tau$	1.000	0.992	-0.008	0.068	0.069
Eq.3 Constant	0.000	-0.006	-0.006	0.081	0.081
Eq.3 Y11.1	0.000	0.001	0.001	0.060	0.060
Eq.3 Y12.1	0.000	-0.000	-0.000	0.061	0.061
Eq.3 Y2.1	0.000	-0.011	-0.011	0.078	0.079
Eq.1 Constant	0.000	-0.003	-0.003	0.065	0.065
Eq.1 Y11.1	0.500	0.488	-0.012	0.054	0.056
Eq.1 Y12.1	0.000	0.002	0.002	0.057	0.057
Eq.1 Y2.1	0.000	0.001	0.001	0.072	0.072
Eq.2 Constant	0.000	0.002	0.002	0.063	0.063
Eq.2 Y11.1	0.000	0.001	0.001	0.056	0.056
Eq.2 Y12.1	0.500	0.492	-0.008	0.054	0.055
Eq.2 Y2.1	0.000	0.000	0.000	0.070	0.070
$\delta_1$	0.000	0.002	0.002	0.071	0.071
$\delta_2$	0.000	-0.002	-0.002	0.070	0.070
Ch.11	1.000	0.988	-0.012	0.043	0.044
Ch.21	0.000	-0.001	-0.001	0.064	0.064
Ch.22	1.000	0.986	-0.014	0.044	0.046

Table 5: Moments of sampling distribution of ML estimator of the parameters CSVAR(1), under DGP1.  $T = 250$ , 1000 MC replications. Parameter names described in Table 2.

		$H_0 : \text{KSVAR}, H_1 : \text{CKSVAR}$			$H_0 : \text{CSVAR}, H_1 : \text{CKSVAR}$		
Sign. Level		10%	5%	1%	10%	5%	1%
DGP1	asymptotic	0.173	0.093	0.020	0.155	0.084	0.022
	bootstrap	0.107	0.045	0.005	0.109	0.052	0.012
DGP2	asymptotic	0.149	0.080	0.018	<b>0.319</b>	<b>0.206</b>	<b>0.073</b>
	bootstrap	0.117	0.050	0.011	<b>0.242</b>	<b>0.141</b>	<b>0.041</b>
DGP3	asymptotic	<b>0.587</b>	<b>0.454</b>	<b>0.244</b>	0.141	0.067	0.019
	bootstrap	<b>0.471</b>	<b>0.365</b>	<b>0.194</b>	0.103	0.053	0.018

Table 6: Rejection frequencies of LR tests of  $H_0$  against  $H_1$  across different DGPs at various significance levels. Computed using 1000 Monte Carlo replications,  $T = 250$ . The asymptotic tests use  $\chi^2_3$  and  $\chi^2_5$  critical values for KSVAR and CSVAR resp. The bootstrap rej. frequencies were computed using the warp-speed method of Giacomini et al. (2013). Bold numbers indicate that the rejection frequencies were computed under  $H_1$  (power).



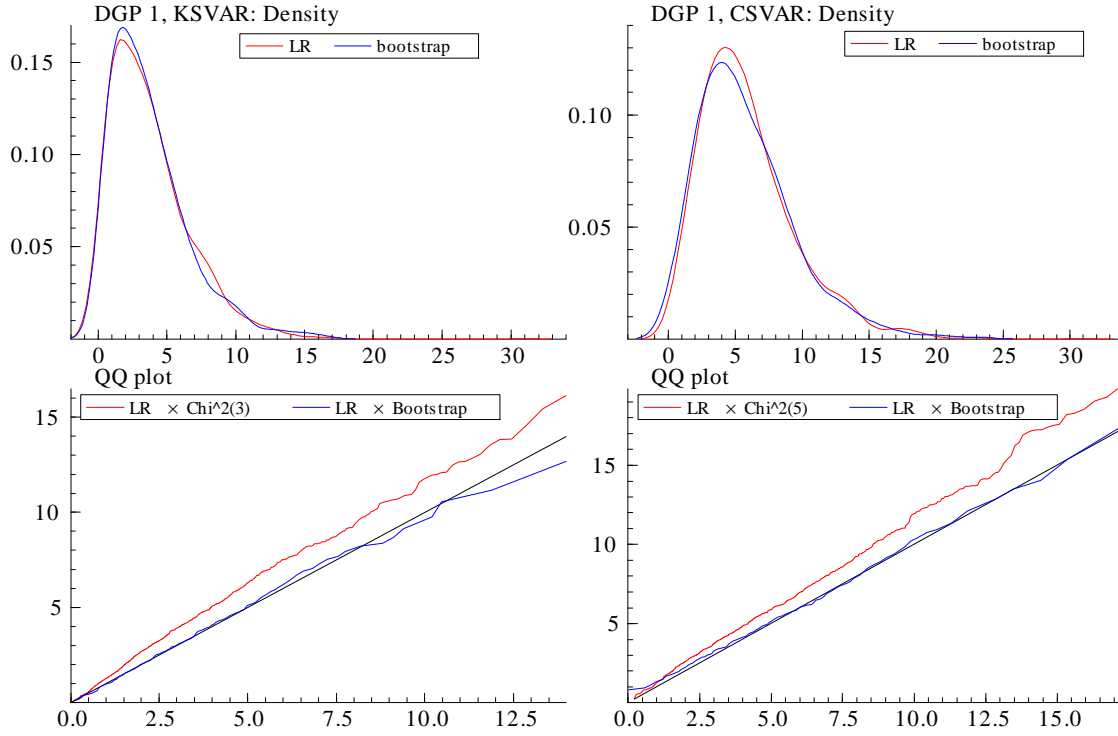


Figure 8: Sampling densities and QQ plots of LR statistics of KSVAR against CKSVAR (left) or CSVAR against CKSVAR (right) in red. Bootstrap densities in blue. Computed for  $T = 250$  using 1000 Monte Carlo replications.

sample gets large, but this level of overrejection is unsatisfactory at  $T = 250$ , which is a typical sample size in macroeconomic applications. The parametric bootstrap appears to do a remarkably good job at correcting the size of the tests. In all cases considered, the parametric bootstrap rejection frequency is not significantly different from the nominal level when the null hypothesis holds (all but the numbers in bold in the Table). To shed further light on this issue, Figures 8 and 9 report the sampling distributions of the two LR statistics and their parametric bootstrap approximations. The sampling distributions of the LR statistics stochastically dominate their asymptotic approximations, but the bootstrap approximations are quite accurate.

Finally, the rejection frequencies highlighted in bold in Table 6 correspond to the power of the tests against two very similar deviations from the null hypothesis. The numbers on the left under DGP3 show the power of the test to reject the KSVAR specification under the alternative at which the coefficient on the latent lag  $B_{2,1}^* = 0.5$ . Similarly, the bold numbers on the right give the power of rejecting CSVAR against the alternative where the coefficient on the observed lag  $B_{22,1} = 0.5$ . Since the lower bound is set to zero, and the sample contains about 50% of observations at the ZLB, the two deviations from the null are of equal magnitude. Yet, we notice the LR test is significantly more powerful against the KSVAR than against the CSVAR. This could be because CSVAR imposes more restrictions than the KSVAR, so one would expect it to have lower power than the KSVAR against similar deviations from the null.

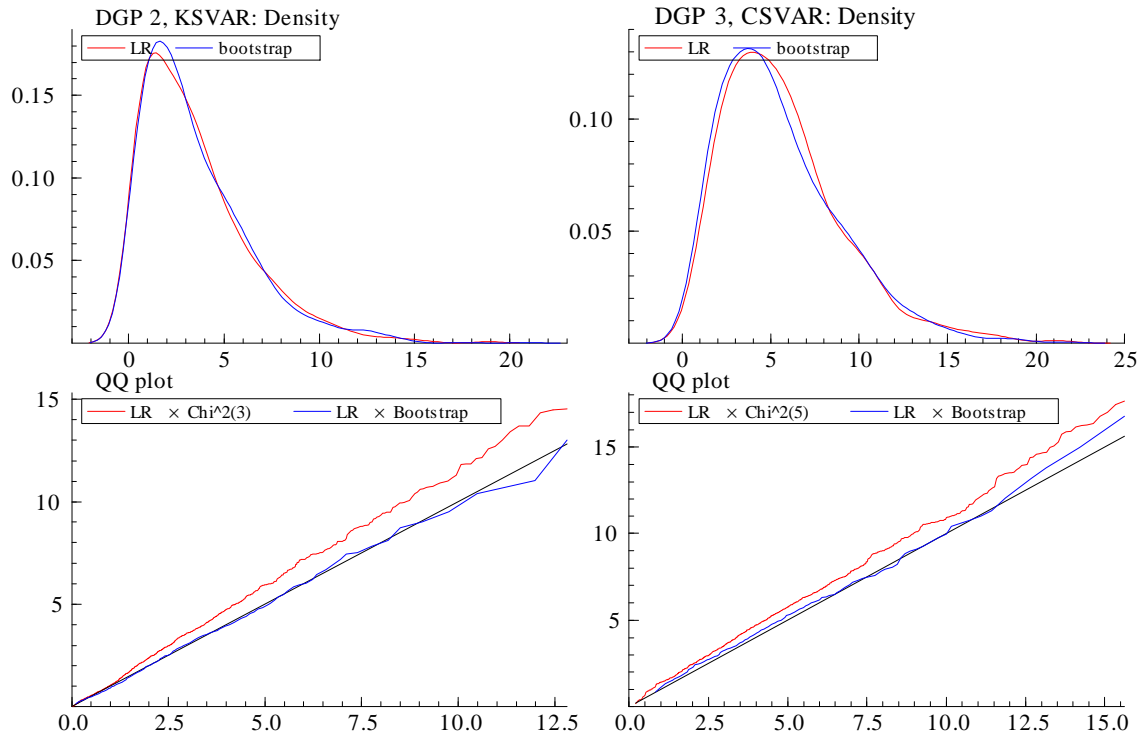


Figure 9: Sampling densities and QQ plots of LR statistics of KSVAR against CKSVAR (left) or CSVAR against CKSVAR (right) in red. Bootstrap densities in blue. Computed for  $T = 250$  using 1000 Monte Carlo replications.

## References

- Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica* 42(6), 999–1012.
- Aruoba, S. B., P. Cuba-Borda, K. Higa-Flores, F. Schorfheide, and S. Villalvazo (2020). Piecewise-Linear Approximations and Filtering for DSGE Models with Occasionally Binding Constraints. Technical report.
- Aruoba, S. B., P. Cuba-Borda, and F. Schorfheide (2017). Macroeconomic dynamics near the ZLB: A tale of two countries. *The Review of Economic Studies* 85(1), 87–118.
- Aruoba, S. B., F. Schorfheide, and S. Villalvazo (2020). SVARs with Occasionally-Binding Constraints. Technical report.
- Ascari, G. and S. Mavroeidis (2020). The unbearable lightness of equilibria in a low interest rate environment. arXiv preprint, arXiv:2006.12966. Available at <https://arxiv.org/abs/2006.12966>.
- Blundell, R. and R. J. Smith (1994). Coherency and estimation in simultaneous models with censored or qualitative dependent variables. *Journal of Econometrics* 64(1-2), 355 – 373.
- Blundell, R. W. and R. J. Smith (1989). Estimation in a class of simultaneous equation limited dependent variable models. *The Review of Economic Studies* 56(1), 37–57.
- Chen, H., V. Cúrdia, and A. Ferrero (2012, November). The macroeconomic effects of large-scale asset

- purchase programmes. *The Economic Journal* 122, F289–F315.
- Clarida, R., J. Galí, and M. Gertler (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics* 115, 147–180.
- Cogley, T. and A. M. Sbordone (2008). Trend inflation and inflation persistence in the New Keynesian Phillips Curve. *American Economic Review* 98(5), 2101–26.
- de Jong, R. and A. M. Herrera (2011). Dynamic censored regression and the open market desk reaction function. *Journal of Business & Economic Statistics* 29(2), 228–237.
- Debortoli, D., J. Gali, and L. Gambetti (2019). On the Empirical (Ir) Relevance of the Zero Lower Bound Constraint. In *NBER Macroeconomics Annual 2019, volume 34*. University of Chicago Press.
- Eggertsson, G. B. and M. Woodford (2003). Zero bound on interest rates and optimal monetary policy. *Brookings papers on economic activity* 2003(1), 139–233.
- Fernández-Villaverde, J., G. Gordon, P. Guerrón-Quintana, and J. F. Rubio-Ramírez (2015). Nonlinear adventures at the zero lower bound. *Journal of Economic Dynamics and Control* 57, 182–204.
- Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics* 7(1), 44–76.
- Giacomini, R., D. N. Politis, and H. White (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric theory* 29(3), 567–589.
- Gourieroux, C., J. Laffont, and A. Monfort (1980). Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes. *Econometrica* 48(3), 675–695.
- Greene, W. H. (1993). *Econometric Analysis*. New York: MacMillan.
- Guerrieri, L. and M. Iacoviello (2015). OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily. *Journal of Monetary Economics* 70, 22–38.
- Hayashi, F. and J. Koeda (2019). Exiting from Quantitative Easing. *Quantitative Economics* 10, 1069–1107.
- Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica* 46(4), 931–959.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153–161.
- Herbst, E. P. and F. Schorfheide (2015). *Bayesian estimation of DSGE models*. Princeton and Oxford: 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. Discussion paper 2020-E-10, Institute for Monetary and Economic Studies, Bank of Japan. Available at <https://www.imes.boj.or.jp/research/papers/english/20-E-10.pdf>.
- Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics* 74(1), 119–147.
- Kulish, M., J. Morley, and T. Robinson (2017). Estimating DSGE models with zero interest rate

- policy. *Journal of Monetary Economics* 88(C), 35–49.
- Lee, L.-F. (1976). Multivariate regression and simultaneous equations models with some dependent variables truncated. "Discussion paper" 76-79, "University of Minnesota", "Minneapolis, USA".
- Lee, L.-F. (1999). Estimation of dynamic and ARCH Tobit models. *Journal of Econometrics* 92(2), 355–390.
- Lewbel, A. (2007). Coherency and completeness of structural models containing a dummy endogenous variable\*. *International Economic Review* 48(4), 1379–1392.
- Liu, P., K. Theodoridis, H. Mumtaz, and F. Zanetti (2019). Changing macroeconomic dynamics at the zero lower bound. *Journal of Business & Economic Statistics* 37(3), 391–404.
- Lütkepohl, H. (1996). *Handbook of Matrices*. England: Wiley.
- Magnusson, L. M. and S. Mavroeidis (2014). Identification using stability restrictions. *Econometrica* 82(5), 1799–1851.
- Malik, S. and M. K. Pitt (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics* 165(2), 190–209.
- Nelson, F. and L. Olson (1978). Specification and estimation of a simultaneous-equation model with limited dependent variables. *International Economic Review* 19(3), 695–709.
- Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics*, Volume 4, pp. 2111–2245. North-Holland.
- Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association* 94(446), 590–599.
- Reifschneider, D. and J. C. Williams (2000). Three lessons for monetary policy in a low-inflation era. *Journal of Money, Credit and Banking* 32(4), 936–966.
- Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics* 85(4), 777–792.
- Rossi, B. (2019). Identifying and estimating the effects of unconventional monetary policy: How to do it and what have we learned? Discussion Paper DP14064, CEPR.
- Smith, R. J. and R. W. Blundell (1986). An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica* 54(3), 679–685.
- Stock, J. H. and M. W. Watson (2001). Vector autoregressions. *Journal of Economic Perspectives* 15(4), 101–115.
- Swanson, E. T. and J. C. Williams (2014). Measuring the effect of the zero lower bound on medium- and longer-term interest rates. *American Economic Review* 104(10), 3154–3185.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26(1), 24–36.
- 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.

Wu, J. C. and J. Zhang (2019). A shadow rate New Keynesian model. *Journal of Economic Dynamics and Control* 107, 103728.