Identification at the Zero Lower Bound

Sophocles Mavroeidis*
University of Oxford and INET

March 14, 2019

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'. Application of the method to US monetary policy using a three-equation SVAR model in inflation, unemployment and the federal funds rate provides some evidence that unconventional policies are partially effective.

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.

The intuition is as follows. By providing an exogenous constraint on policy, the ZLB acts like a 'quasi experiment' introducing random variation to policy that can be used to identify the monetary policy shock. Once this is identified, the entire impulse response function can be obtained. This idea is potentially generalizable to other models with occasionally binding constraints.

*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, and the DNB for useful comments and discussion, as well as Mishel Ghassibe, Lukas Freund, Shangshang Li, and Patrick Vu for research assistance.

There are similarities between identification via occasionally binding constraints and identification via structural change proposed by Magnusson and Mavroeidis (2014). Structural change induces different regimes, and the switch from one regime to another generates variation that identifies parameters that are constant across regimes. For example, an exogenous shift in a policy reaction function identifies the transmission mechanism, provided the latter is unaffected by the policy shift.[1] An instrumental variables interpretation of this 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.[2] 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 constructively in structural vector autoregressions (SVARs).

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).

The second 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. It is not possible to impose the ZLB constraint using Markov switching models, as was done in Hayashi and Koeda (2013, 2014), because Markov-switching cannot guarantee that the constraint will be respected with probability one, and also typically does not account for the fact that the switch from one regime to the other depends on the structural shocks. 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,

[1] 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).

[2] Magnusson and Mavroeidis (2014) deal with the additional complication that the timing and number of regimes might be unknown.

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. 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 to implement any of the existing identification schemes.

Identification of the causal effects of policy from the ZLB relies on two conditions. The first condition is that the ZLB "does represent an important constraint on what monetary policy can achieve" as argued by Eggertsson et al. (2003). This condition rules out the following scenario. Suppose the conventional policy instrument is subject to a lower bound, but policy makers can still achieve their policy objectives fully using "unconventional" policies. Such a scenario could be characterized by a model in which the "effective" policy instrument is unconstrained but we only observe a censored version of it. In this case, there is no actual regime change when the observed policy instrument hits the constraint, and so there is no additional information to identify the causal effects of policy. In the opposite case, i.e., when unconventional policies are completely ineffective, we obtain point identification of the policy effects. Finally, in intermediate cases, i.e., when unconventional policy is only partially effective, I show that the policy effects are only partially (set) identified. The methodology I develop covers all of the above cases. It also provides a simple test of the efficacy of unconventional policy.

The second condition for identification of the policy effects via the ZLB is that the parameters of the transmission mechanism are the same across the different regimes induced by the ZLB. In general, this condition seems fairly innocuous, since the parameters of the transmission mechanism typically depend on preferences and technology, and there is no obvious reason to expect these to be affected by whether policy is constrained or not. Structural models that are immune to the well-known Lucas (1976) critique satisfy this assumption. In micro-founded dynamic stochastic general equilibrium (DSGE) models this assumption is seemingly uncontroversial and it is commonly imposed in the literature, see, e.g., Aruoba et al. (2016); Mertens and Ravn (2014); Fernández-Villaverde et al. (2015) and the references therein. However, in SVARs, this assumption is potentially more controversial. This is because with occasionally binding constraints there is no piecewise linear SVAR representation of a generic DSGE model under rational expectations (there could be under backward-looking or naive expectations), as the solutions of those models are typically highly nonlinear, see, e.g., Fernández-Villaverde et al. (2015), Guerrieri and Iacoviello (2015), and Aruoba et al. (2017). Therefore, if one has a strong prior on a micro-founded rational expectations DSGE representation, a piecewise linear SVAR model would be misspecified and should not be used for causal inference.

However, the methodology developed in this paper still provides a valid test of the null hypothesis that unconventional policy is fully effective, i.e., that the ZLB is empirically irrelevant, as was argued recently by Debortoli et al. (2018) and Swanson (2018). It also provides a valid method to analyze a linear SVAR with conventional identifying restrictions if the aforementioned hypothesis is not rejected.

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 the constancy or otherwise of the reaction function 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 be consistent with unconventional policy being less active (during ZLB regimes) than conventional policy (during non-ZLB regimes), i.e., with policy objectives being different across regimes. This is a fundamental identification problem that is difficult to overcome without additional information. The conclusions in the empirical analysis below on the efficacy of unconventional policies are drawn under the assumption that there is no difference in the policy objectives, and hence the policy reaction function, across regimes. This seems to be a reasonable starting point for an initial analysis of the data, especially since this is mainly intended as an illustration of the methods developed in the paper. A more thorough investigation of this issue should incorporate additional information, such as other 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.

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 link to the related microeconometrics literature.

Consider a system of simultaneous equations in two endogenous variables $y = (y_1, y_2)'$, with $y_2 \geq b$. For example, y_2 can be interpreted as the policy instrument and y_1 is some target macroeconomic variable of interest. The model is given by the equations

$$y_1 = \beta y_2 + \alpha_1' x + \varepsilon_1 \tag{1}$$

$$y_2 = \max\left\{\gamma y_1 + \alpha_2' x + \varepsilon_2, b\right\} \tag{2}$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ are the structural shocks and x are exogenous variables, such that $E(\varepsilon|x) = 0$. Assume that $\varepsilon \sim N(0, \Sigma)$, with $\Sigma = diag\left(\sigma_1^2, \sigma_2^2\right)$. The unknown structural parameters are $\theta = \left(\beta, \gamma, \alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2\right)'$, and b is assumed to be known (but does not need to be constant over the sample). I will refer to the model given by equations (1) and (2) as a kinked simultaneous equations model (KSEM). KSEM reduces to a standard linear SEM when $b = -\infty$.

A variant of the above model was first studied in Amemiya (1974) and Lee (1976), as a multivariate extension of the well-known Tobit model (Tobin, 1958).[3] The first reference I found to the model (1) and (2) is in (Nelson and Olson, 1978, eq. 6 and 7). Nelson and Olson (1978) wrote it down as an example of a model that was less suitable for microeconometric applications than the following alternative specification, which I will refer to as a censored simultaneous equation model (CSEM):

$$y_1 = \beta y_2^* + \alpha_1' x + \varepsilon_1, \tag{3}$$

$$y_2^* = \gamma y_1 + \alpha_2' x + \varepsilon_2, \tag{4}$$

$$y_2 = \max\left\{y_2^*, b\right\}, \tag{5}$$

where y_2^* is a latent variable. Nelson and Olson (1978) focused their analysis only on the CSEM, which subsequently became the main focus of the literature (Smith and Blundell, 1986; Blundell and Smith, 1989). The key distinction between the KSEM (1)-(2) and CSEM (3)-(5) is that in the former, y_1 depends on the *observed* y_2, while in the latter it depends on the latent variable y_2^* which is only partially observed due to censoring. This distinction is crucial for identification.

2.1 Coherency and completeness

It is important to point out an issue that arises in (1)-(2) but not in (3)-(5): the likelihood does not exist without restrictions on the parameter space. Specifically, there are values of θ and ε for which there are either no solutions or multiple solutions of (1)-(2) for y. Amemiya (1974) refers to these as "type 1 and type 2 difficulties", respectively, also known as 'incoherency' and 'incompleteness' in the microeconometrics literature (Blundell and Smith, 1994). Coherency and completeness of a structural model together imply existence and uniqueness of a reduced form (Lewbel, 2007). Amemiya (1974) notes that the problem is akin to the Complementarity Problem in the programming literature, where necessary and sufficient conditions can be found. In the present example, these are $1 - \gamma\beta > 0$ (Nelson

[3] The difference is that in the models of Amemiya (1974) and Lee (1976), all endogenous variables are truncated.

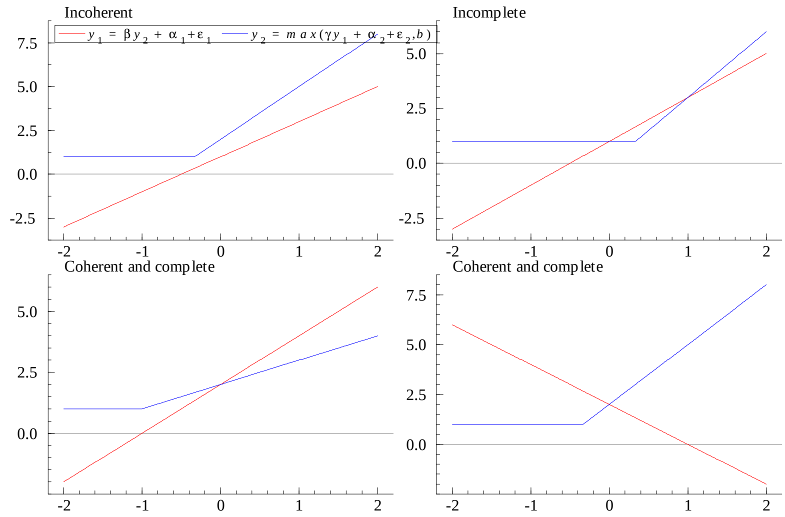

Figure 1: Coherency and completeness of the Kinked Simultaneous Equations model.

and Olson, 1978). This issue is illustrated in Figure 1. Specifically, coherency and completeness are satisfied in this example when the slope of equation (2) is strictly smaller than that of equation (1).

Despite its importance, the coherency condition (which is the same as existence of an equilibrium) is rarely addressed in the literature on the ZLB. This is probably because in nonlinear DSGE models it is analytically intractable and also hard to characterize numerically. However, I will show in the next section that in SVARs the condition is straightforward to characterize analytically and check in any given application. The condition is essential for estimation, and also provides additional restrictions on the admissible range of the structural parameters.

2.2 Nesting KSEM and CSEM models

The KSEM and CSEM models can be nested by combining equations (1) and (3) into the following equation:

$$y_1 = \beta_2^* y_2^* + \beta_2 y_2 + \alpha_1' x + \varepsilon_1, \tag{6}$$

The model given by equations (6), (4) and (5) can be termed censored and kinked simultaneous equations model, or CKSEM.

Reparametrize β_2^* and β_2 into $\beta_2^* = \lambda\beta$ and $\beta_2 = (1-\lambda)\beta$. Then, equation (6) can be equivalently written as

$$y_1 = \beta\left(\lambda y_2^* + (1-\lambda) y_2\right) + \alpha_1' x + \varepsilon_1. \tag{7}$$

If (4) gives the desired policy stance in the absence of any constraints, and there exist alternative policies not modelled explicitly that can partially mitigate the effect of the constraint, then the variable $y_s := \lambda y_2^* + (1-\lambda) y_2$ can be thought of as a measure of the *effective* policy stance. When $\lambda = 1$, policy is completely unconstrained, and the CKSEM reduces to a CSEM, while if $\lambda = 0$, unconventional policy is completely ineffective and the model reduces to the KSEM. Hence, λ is interpretable as a measure of the efficacy of unconventional policy.

2.3 Identification

The CSEM (3)-(5) is an unrestricted linear SEM in the variables (y_1, y_2^*), so it is underidentified. Censoring is simply a measurement issue that complicates estimation and inference but does not affect identification of the structural parameters. I will now show that the CKSEM, given by equations (7), (4) and (5), is identified when $\lambda \neq 1$. Specifically, it is point-identified when λ is known, e.g., $\lambda = 0$ in the case of the KSEM (1)-(2), and partially identified when λ is unknown. This result appears to be new in the literature and is, in my view, the most interesting contribution of the paper.

Substituting for y_2^* in (7) using (4), and rearranging, we obtain

$$y_1 = \widetilde{\beta} y_2 + \widetilde{\alpha}_1' x + \widetilde{\varepsilon}_1, \text{ where} \tag{8}$$

$$\widetilde{\beta} = \frac{(1-\lambda)\beta}{1-\lambda\beta\gamma}, \widetilde{\alpha}_1 = \frac{\alpha_1 + \lambda\beta\alpha_2}{1-\lambda\beta\gamma}, \widetilde{\varepsilon}_1 = \frac{\varepsilon_1 + \lambda\beta\varepsilon_2}{1-\lambda\beta\gamma}. \tag{9}$$

Note that, since equation (8) is isomorphic to (1), the coherency and completeness condition for the CKSEM is analogous to that of the KSEM (1)-(2), i.e., $\gamma\widetilde{\beta} < 1$, which, using (9), is equivalent to $(1-\gamma\beta)/(1-\lambda\gamma\beta) > 0$. Similarly, the discussion of identification of (8) is analogous to that of the KSEM. Therefore, I discuss first the identification of the KSEM ($\lambda = 0$) and then turn to the general case.

2.3.1 Identification of KSEM

Let $D := 1_{\{y_2 = b\}}$ be an indicator (dummy) variable that takes the value 1 when y_2 is on the boundary and zero otherwise. When $\gamma\beta < 1$ (coherency), the reduced form of the KSEM is given by the equations

$$y_1 = \mu_1' x + u_1 + \beta D \left(b - \mu_2' x - u_2\right), \text{ and} \tag{10}$$

$$y_2 = \max\left(\mu_2' x + u_2, b\right), \tag{11}$$

where

$$u_1 = \frac{\varepsilon_1 + \beta\varepsilon_2}{1-\gamma\beta}, \quad u_2 = \frac{\gamma\varepsilon_1 + \varepsilon_2}{1-\gamma\beta}, \tag{12}$$

$$\mu_1 = \frac{\alpha_1 + \beta\alpha_2}{1-\gamma\beta}, \quad \mu_2 = \frac{\gamma\alpha_1 + \alpha_2}{1-\gamma\beta}.$$

Equation (11) is a standard univariate limited dependent variable model. With Gaussian errors, it is exactly a Tobit model (Tobin, 1958). Its parameters, μ_2 and $\tau^2 := var(u_2)$, are therefore identified in the usual way by Tobit regression. Equation (10) is an 'incidentally kinked' regression model. It is similar to an incidentally censored regression, and, with Gaussian errors, its parameters are identified by a variant of the well-known Heckit method (Heckman, 1979). The details of this, as well as various alternative estimators, are derived in the next section for the full SVAR model.

Here, it is instructive to provide an alternative instrumental variables interpretation of identification. Consider using the regime indicator $D = 1_{\{y_2=b\}}$ as an instrument for the endogenous regressor y_2 in the structural equation (1). The corresponding IV estimator identifies

$$\beta_{IV} := \frac{E(y_1|x, D=1) - E(y_1|x, D=0)}{E(y_2|x, D=1) - E(y_2|x, D=0)},$$

which is inconsistent for β because D is, in fact, endogenous:

$$\beta_{IV} = \beta + \frac{E(\varepsilon_1|x, D=1) - E(\varepsilon_1|x, D=0)}{E(y_2|x, D=1) - E(y_2|x, D=0)} \neq \beta. \tag{13}$$

However, if we know the distribution of the reduced-form errors, we can compute the bias of the IV estimator, given by the second term in the right-hand side of (13), and point-identify β. A simple way to do this is via the control function approach (Heckman et al., 1978).[4] Let

$$h(\mu_2, \tau) := (1-D)\left(y_2 - \alpha_2' x\right) - D\frac{\tau\phi(a)}{\Phi(a)}, \quad a := \frac{b - \mu_2' x}{\tau},$$

and $\phi(\cdot)$, $\Phi(\cdot)$ are standard Normal density and distribution functions, respectively. The control function $h(\mu_2, \tau)$ captures the endogeneity in y_2 in equation (1), i.e., $E(\varepsilon_1|y_2, x) = \rho h(\mu_2, \tau)$, for some constant ρ.[5] Hence, β can be identified from the regression

$$E(y_1|y_2, x) = \beta y_2 + \mu_1' x + \rho h(\mu_2, \tau). \tag{14}$$

The rank condition for the identification of β is simply that the regressors are not perfectly collinear. This holds if and only if $0 < \Pr(D=1) < 1$. So, as long as there are some but not all the observations at the boundary, the model is generically identified.

Importantly, the control function only depends on the parameters in one of the two regimes, when $D = 0$. So, it does not matter whether the policy rule (4) changes when y_2 hits the lower bound.

2.3.2 Partial identification of the CKSEM

The discussion of the previous subsection carries over to the CKSEM given by equations (7), (4) and (5), when β, α_1 and ε_1 in (10)-(12) are replaced by $\widetilde{\beta}, \widetilde{\alpha}_1$, and $\widetilde{\varepsilon}_1$, defined in (9). Therefore, $\widetilde{\beta}$ is identified as the coefficient on the kink in the reduced-form equation for y_1 (10). From (9), it follows

[4] I am grateful to Andros Kourtelos for that suggestion.

[5] Specifically, $\rho = cov(\varepsilon_1, u_2)/\tau^2 - \beta$.

that β, α_1 and ε_1 are underidentified unless λ is known.

Now, consider the restriction $\lambda \in [0, 1]$. Given the interpretation of λ 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 unconventional policy on y_1 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 β is completely unidentified when $\lambda = 1$ (which corresponds to the CSEM). From the definition of $\widetilde{\beta}$ in (9) and the fact that $1 - \lambda\beta\gamma \neq 0$ from the coherency condition, it follows that $\lambda = 1$ implies $\widetilde{\beta} = 0$. So, when $\widetilde{\beta} = 0$, β is completely unidentified. It remains to see what happens when $\widetilde{\beta} \neq 0$. 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 $\gamma_0 := \omega_{12}/\omega_{11}$. It is also the coefficient of the regression of y_2 on y_1 over the uncensored observations. This can be interpreted as the value the reaction function coefficient γ in (4) would take if $\beta = 0$, i.e., the value corresponding to a Choleski identification scheme where y_2 is placed last. γ_0 is obviously identifiable from the reduced form. In the Appendix, I prove the following bounds

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

In words, when the coefficient on the kink in the reduced form, $\widetilde{\beta}$, is different from zero, and it is of the same sign as the coefficient of the regression of y_2 on y_1 over the uncensored observations, γ_0, and if $\widetilde{\beta}\gamma_0 \leq 1$, then we can identify both the sign of the causal effect of y_2 on y_1 and get bounds on (the absolute value of) its magnitude. In those cases, the identified coefficient $\widetilde{\beta}$ is an attenuated measure of the true causal effect β. Moreover, $\widetilde{\beta}\gamma_0 > 1$ implies that the coefficient on the latent variable y_2^* is of the opposite sign than the coefficient on the observed variable y_2, i.e., unconventional policy has the opposite effect of the conventional one. That could be interpreted as saying that unconventional policy is counterproductive.

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 $\widetilde{\beta} = 0$, so $\widetilde{\beta} \neq 0$ implies $\lambda \neq 1$. Therefore, rejecting $\widetilde{\beta} = 0$ unambiguously implies that the bound is empirically relevant.[6] In fact, the parameter λ is also set-identified. The sharp identified set is as follows:

[6] The converse is not true since $\widetilde{\beta} = 0$ can also arise from $\beta = 0$ even when $\lambda \neq 1$.

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 kth variable Y_{2t} is bounded from below by b. 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 pth-order SVAR model is given by the equations

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

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

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

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

Equation (17) can be interpreted as a policy reaction function, as it determines the desired policy stance Y_{2t}^*. Similarly, ε_{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 exogenous variables, 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.

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

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

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

The vector of structural errors ε_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 CKSEM model, where $Y_{2t}^* < b$ has no (contemporaneous) impact on Y_t. In the dynamic setting, it feels natural to define the corresponding 'kinked SVAR' model (KSVAR) as a model in which Y_{2t}^* has neither a contemporaneous nor a dynamic effect on Y_t. 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 policy at all horizons.

The opposite extreme to the KSVAR is the censored SVAR model (CSVAR). Again, unlike the CSEM, which only deals with contemporaneous effects, the idea of a CSVAR is to impose the assumption that the constraint is ineffective 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, in accordance with the terminology in the previous section, I refer to the general model given by (18) as the 'censored and kinked SVAR' (CKSVAR).

Define the $k \times k$ square matrices

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

$\overline{A}$ 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 if and only if*

$$\kappa := \frac{\overline{A}_{22} - A_{21}A_{11}^{-1}\overline{A}_{12}}{A_{22}^* - A_{21}A_{11}^{-1}A_{12}^*} > 0. \tag{20}$$

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. The proposition follows as a corollary to (Gourieroux et al., 1980, Theorem 2). An alternative proof is given in the Appendix.

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:

$$D_t = 1_{\{Y_{2t}=b\}}. \tag{21}$$

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

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

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

$$\overline{Y}_{2t}^* = \overline{C}_2 X_t + \overline{C}_2^* \overline{X}_t^* + u_{2t}, \tag{24}$$

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

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

$$\widetilde{\beta} = \left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right)^{-1} \left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12} \right), \tag{26}$$

κ *is defined in (20) and the matrices* $\overline{C}_1, \overline{C}_2$*, are given in the Appendix.*

Note that the "reduced-form" latent process $\overline{Y}_{2t}^*$ 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 $\overline{C}_2^* \neq 0$. 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 ψ denote the parameters that characterize the reduced form (22)-(23): $\widetilde{\beta}, \overline{C}$, $\overline{C}^*$ and $\Omega = var(u_t)$. It is useful to decompose ψ into $\psi_2 = \left(\overline{C}_2, \overline{C}_2^*, \tau \right)'$, where $\tau = \sqrt{var(u_{2t})}$, and $\psi_1 = \left(vec\left(\overline{C}_1\right)', vec\left(\overline{C}_1^*\right)', \widetilde{\beta}', \delta', vech\left(\Omega_{1.2}\right) \right)'$, where $\delta = \Omega_{12}/\tau^2$, $\Omega_{1.2} = \Omega_{11} - \delta\delta'\tau^2$, and $\Omega_{ij} = cov(u_{it}, u_{jt})$.

Equation (23) is the dynamic Tobit regression model studied by Lee (1999). So, its parameters, ψ_2, are generically identified provided that the regressors are not perfectly collinear. This requires that $0 < \Pr(D_t = 1) < 1$.

Given ψ_2, the identification of the remaining parameters, ψ_1, can be characterized using a control function approach. Consider the $k-1$ regression equations

$$E\left(Y_{1t} | Y_{2t}, X_t, \overline{X}_t^* \right) = \overline{C}_1 X_t + \overline{C}_1^* \overline{X}_t^* + \widetilde{\beta} Z_{1t} + \delta Z_{2t}, \tag{27}$$

where

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

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

$a_t = \left(\frac{b - \overline{C}_2 X_t - \overline{C}_2^* \overline{X}_t^*}{\tau}\right)$, and $\phi(\cdot)$, $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively. When $\overline{C}^*$ is different from zero, regressors $\overline{X}_t^*$, Z_{1t}, and Z_{2t} in (27) are unobserved, so we need to replace them with their expectations conditional on $Y_{2t}, Y_{t-1}, ..., Y_1$. Then, the regressors on the right-hand side of (27) become $\mathbf{X}_t := \left(X_t', \overline{X}_{t|t}^{*\prime}, Z_{1t|t}, Z_{2t|t}\right)'$, where $h_{t|t} := E\left(h\left(\overline{X}_t^*\right)|Y_{2t}, Y_{t-1}, ..., Y_1\right)$ for any function $h(\cdot)$ whose expectation exists.[7] The coefficients $\overline{C}_1, \overline{C}_1^*, \widetilde{\beta}$, and δ 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 ψ is $k_0 k + k^2 p + kp + k(k+1)/2$. The number of structural parameters in (18) is $k_0 k + k^2 p + kp + k^2 + k$. So, the CKSVAR is underidentified by $k(k+1)/2$ restrictions. Nevertheless, I will show that the impulse responses to ε_{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).[8] The IRF to ε_{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 function of ε_{2t} and a process that is orthogonal to it, i.e., $u_t = g(\varepsilon_{2t}, e_t)$, where e_t is independent of ε_{2t}. From Proposition 2, it follows that the g function is linear, and more specifically,

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

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

[7] In the KSVAR model, we have $\overline{C}_1^* = 0$ and $\overline{C}_2^* = 0$, so $\overline{X}_t^*$ drops out of (27), and the regressors Z_{1t}, Z_{2t} are observed, so $Z_{jt|t} = Z_{jt}$, $j = 1, 2$.

[8] For the empirical analysis below, I will use the conditional "generalized impulse response function" defined in (Koop et al., 1996, eq. (3)), see (38).

where

$$\overline{\beta} := -A_{11}^{-1}\overline{A}_{12}, \quad \overline{\gamma} := -\overline{A}_{22}^{-1}A_{21}, \tag{32}$$
$$\bar{\varepsilon}_{1t} := A_{11}^{-1}\varepsilon_{1t}, \quad \bar{\varepsilon}_{2t} := \overline{A}_{22}^{-1}\varepsilon_{2t},$$

and $\overline{A}_{22} = A_{22}^* + A_{22}$, defined in (19). Note that $\overline{\beta}$ can be interpreted as the marginal effect of Y_{2t} on Y_{1t}, and $\overline{\gamma}$ is the marginal effect of Y_{1t} on Y_{2t} (the contemporaneous reaction function coefficients) when $Y_{2t} > b$ (unconstrained regime). The shock vector $\bar{\varepsilon}_{1t}$ is not structural but it is orthogonal to ε_{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 $\overline{\beta}$, $\overline{\gamma}$, and $\overline{A}_{22}$ 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 $\overline{\beta} = \widetilde{\beta}$,*

$$\overline{\gamma} = \left(\Omega_{12}' - \Omega_{22}\overline{\beta}'\right)\left(\Omega_{11} - \Omega_{12}\overline{\beta}'\right)^{-1}, \text{ and} \tag{33}$$
$$\overline{A}_{22}^{-1} = \sqrt{(-\overline{\gamma}, 1)\,\Omega\,(-\overline{\gamma}, 1)'}. \tag{34}$$

Remarks 1. $\overline{\beta} = \widetilde{\beta}$ 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 ε_{2t} and $\bar{\varepsilon}_{1t}$.

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

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

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:

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

where $\bar{B}_2 = \overline{A}_{22}^{-1}B_2$, $\bar{B}_2^* = \overline{A}_{22}^{-1}B_2^*$. 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

$$Y_{2t} = \max\left(\widetilde{\gamma}Z_t + \tilde{B}_2X_t + \tilde{B}_2^*X_t^* + \tilde{\varepsilon}_{2t}, b\right), \tag{36}$$

where $\widetilde{\gamma} = \left(1 - \overline{\gamma}\overline{\beta}\right)^{-1} \overline{\gamma}$, $\tilde{B}_2 = \left(1 - \overline{\gamma}\overline{\beta}\right)^{-1} \bar{B}_2$, $\tilde{B}_2^* = \left(1 - \overline{\gamma}\overline{\beta}\right)^{-1} \bar{B}_2^*$ and $\tilde{\varepsilon}_{2t} = \left(1 - \overline{\gamma}\overline{\beta}\right)^{-1} \bar{\varepsilon}_{2t}$. Note that the coherency condition (20) becomes $\kappa = \frac{\overline{A}_{22}}{A_{22}^*}\left(1 - \overline{\gamma}\overline{\beta}\right) > 0$, so $1 - \overline{\gamma}\overline{\beta} \neq 0$, which guarantees the existence of the representation (36). Given $\overline{\beta} = \widetilde{\beta}$, the structural parameter $\overline{\gamma}$ can then be obtained as $\overline{\gamma} = \widetilde{\gamma}\left(I_{k-1} + \widetilde{\beta}\widetilde{\gamma}\right)^{-1}$, 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 ε_{2t} during periods when $Y_{2t} = b$, is not identified.[9] Similarly, $\kappa = \frac{\overline{A}_{22}}{A_{22}^*}\left(1 - \overline{\gamma}\overline{\beta}\right)$ 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 $\overline{\beta}, \overline{\gamma}$ and $\overline{A}_{22}$, not A_{22}^* or κ. The change allowed for by $A_{22} \neq 0$ is of a specific type.

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 CKSEM model in equation (7). Specifically, if $\overline{A}_{12} = A_{12} + A_{12}^*$ 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 $A_{12}^* = \Lambda\overline{A}_{12}$ and $A_{12} = \left(I_{k-1} - \Lambda\right)\overline{A}_{12}$, where $\Lambda = diag\left(\lambda_j\right)$, $\lambda_j \in [0,1]$ for $j = 1, ..., k-1$. I further impose the restriction that $\lambda_j = \lambda$ for all j, so that $A_{12}^* = \lambda\overline{A}_{12}$ and $A_{12} = (1 - \lambda)\overline{A}_{12}$ 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.

We also need to consider the impact of A_{22} on identification. The variable $\zeta = \overline{A}_{22}/A_{22}^*$ 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 $\overline{A}_{22}^{-1}A_{21}$. 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 $\widetilde{\beta}$ in (9) can be written as

$$\widetilde{\beta} = (1 - \xi)\left(I - \xi\overline{\beta}\overline{\gamma}\right)^{-1}\overline{\beta}, \quad \xi := \lambda\zeta. \tag{37}$$

[9] 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.

Similarly, given $\zeta > 0$, the coherency condition (20) reduces to $\left(1 - \overline{\beta}\overline{\gamma}\right)\left(1 - \xi\overline{\gamma}\overline{\beta}\right) > 0$. Notice that the parameters λ, ζ only appear multiplicatively, so it suffices to consider them together as $\xi = \lambda\zeta$. Once $\overline{\beta}$ is known, the remaining structural parameters needed to obtain the IRF to ε_{2t} are $\overline{\gamma}$ and $\overline{A}_{22}$, and they are obtained from Proposition 3. So, the identified set can be characterized by varying ξ over its admissible range. Without further restrictions on ζ, 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 CKSEM in the previous section where λ played the role of ξ. In the case $k = 2$, the bounds derived in eq. (15) apply, with $\beta = \overline{\beta}$ 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 $\overline{\beta}$. 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 ξ. An algorithm for doing this is provided in Appendix B.3.

3.2 Estimation

Estimation of the KSVAR is straightforward, since the likelihood is analytically available under Gaussian errors. 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 of the presence of latent lags and the likelihood function of the reduced form given in Proposition 2 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 method of moments (MM), and sequential estimation motivated by the constructive identification proof in Section 3. These are also described in the 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 χ^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 negative autoregressive roots that are larger than 1 in absolute value do 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 and related frequentist 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 various estimators discussed above. 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 important 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) will 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).[10] The variables are plotted in Figure 2 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 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.

[10] The inflation data are computed as $\pi_t = 400 1n(P_t/P_{t-1})$, 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.

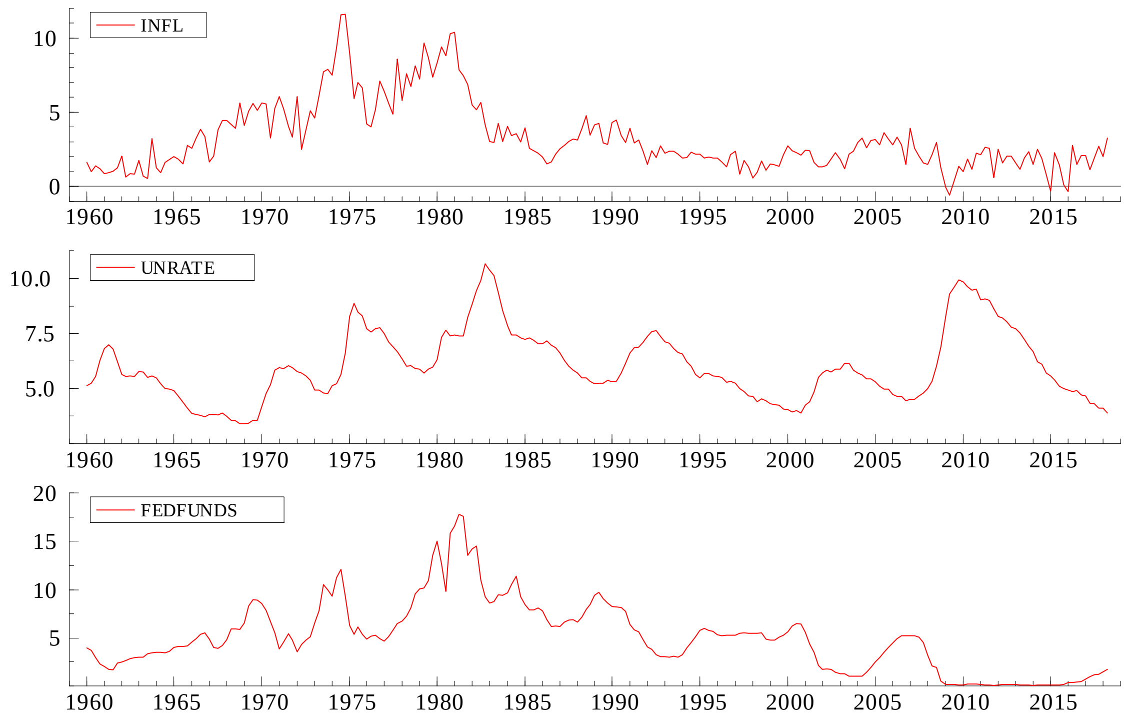

Figure 2: Data used in Stock and Watson (2001) over the extended sample 1960q1 to 2018q2.

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	0.015
CSVAR(4)	-94.86	-94.87	14	26.43	0.023	0.129

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 is unrestrcted 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 χ^2_q, q = number of restrictions. Bootstrap p-values from parametric bootstrap with 999 replications.

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 $\overline{C}_1^* = 0$ and $\overline{C}_2^* = 0$ 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 $\overline{C}^* = \left(\overline{C}_1^{*\prime}, \overline{C}_2^{*\prime}\right)' \neq 0$ would imply that unconventional policy has at least a lagged effect on Y_{1t}. $\overline{C}^* = 0$ 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, meaning that unconventional policy is at least partially effective.

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 C_1 and C_2 corresponding to lags of Y_{2t} in (22) and (23) are all zero. In addition, it imposes the restriction that $\widetilde{\beta} = 0$ in (22), that is, 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 at the 5% level, but the bootstrap p-value is 0.129. 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 IRF 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):

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

If t is such that $\overline{Y}_{2,t-s}^{*} = Y_{2,t-s} > b$ for all $s = 1, ..., p$, then $\overline{X}_t^{*} = 0$,[11] and hence $IRF_{h,t}(\varsigma, X_t, 0)$ is observed. However, if t is such that any element of $\overline{X}_t^{*}$ is unobserved, we could either evaluate $IRF_{h,t}(\varsigma, X_t, \cdot)$ at an estimated value of $\overline{X}_t^{*}$ conditional on the observed data, or we can estimate $IRF_{h,t}\left(\varsigma, X_t, \overline{X}_t^{*}\right)$ conditional on the observed data. When we estimate the IRF from a CKSVAR(4) at the end of our sample, $t = 2018q3$, $\overline{X}_t^{*} = 0$ because the Federal funds rate was above the ZLB over the previous four quarters, so filtering is unnecessary.

Figure 3 reports the nonrecursive IRFs to a 100 basis points monetary policy shock from the CKSVAR under the assumption that $\lambda = 0$ (unconventional policy is ineffective) and two different

[11] Recall the defintion $\overline{X}_t^{*} = (\overline{x}_{t-1}, ..., \overline{x}_{t-p})'$, $\overline{x}_t = \min\left(\overline{Y}_{2t}^{*} - b, 0\right)$, given in Proposition 2.

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, we turn to the identified sets of the IRFs that arise when we relax the restriction that unconventional policy is ineffective, i.e., λ can be greater than zero. We consider the range of $\xi = \lambda\zeta \in [0, 1]$, recalling that λ measures the efficacy of unconventional policy and ζ 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 4 report the identified sets without any other restrictions. The right-hand-side graphs derive the identified sets when we 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 3.

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 $\overline{\beta}, \overline{\gamma}$ 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

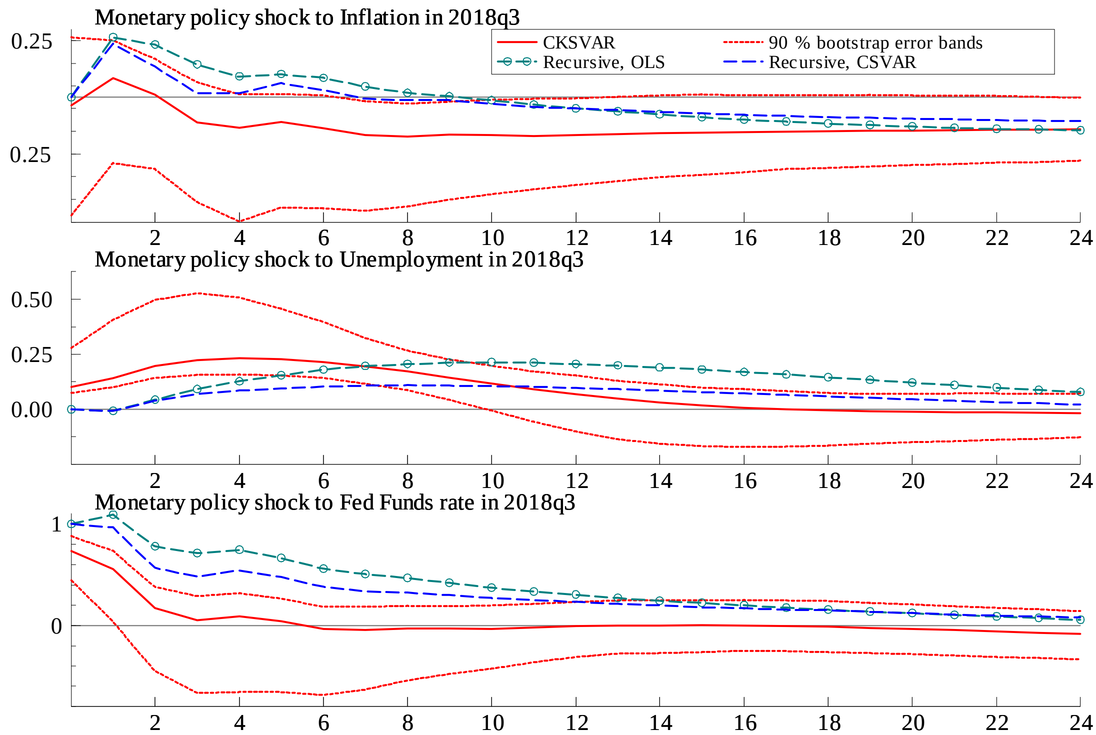

Figure 3: IRFs to a monetary policy shock from a three equation CKSVAR(4) estimated over the period 1960q1 to 2018q2. The solid red line corresponds to the nonrecursive identification from the ZLB under the assumption that unconventional policy is ineffective. The dashed blue line corresponds to the nonlinear recursive IRF, estimated with the CKSVAR(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.

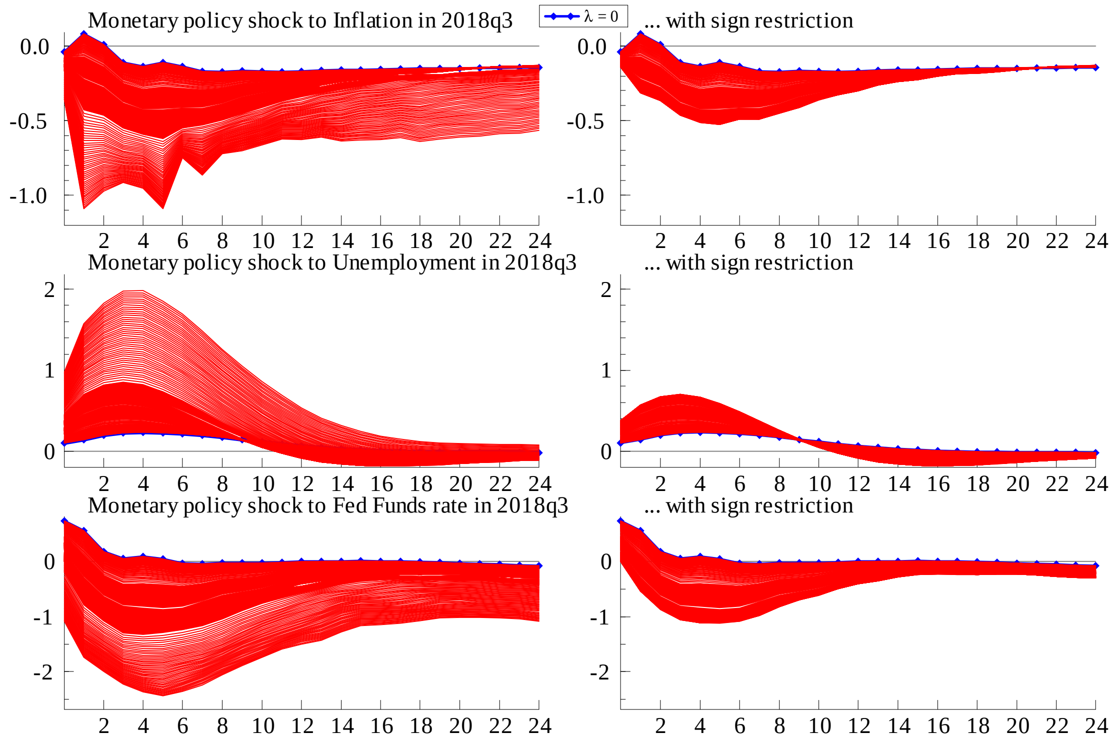

Figure 4: Identified sets of the IRFs to a 100bp monetary policy shock in a CKSVAR(4) in 2018q3. The red area denotes the identified set, the blue line with diamonds indicates the point identified IRF under the restriction that unconventional policy is ineffective $(\lambda = 0)$. The figures on the left provide the full identified set. The figures on the right impose the restirction that the response of the Fed funds rate be nonnegative.

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 ζ, the method can be used to obtain an estimate of the identified set for λ, 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 λ is $[0.0.506]$. In other words, the identified set excludes values of the efficacy of policy beyond 51%, so that, roughly speaking, unconventional policy is at most 50% as effective as conventional one. Note that this estimate does not account for sampling uncertainty and relies crucially on the assumption about the constancy of the reaction function across the two regimes. As argued in the introduction, this assumption can be motivated by arguing that there is no reason to believe that policy objectives (and hence the reaction function) may have shifted over the ZLB period. However, to illustrate the implications of relaxing that assumption, consider the following alternative assumption. Suppose $\zeta = 1/2$, 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 λ would include 1, i.e., the data would be consistent with the view that unconventional policy is fully effective. So, under this assumption, the reason we observed a subdued response to policy shocks over the ZLB period is because policy was less active over that period.

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 we will require additional information to untangle the structural parameters λ and ζ from $\xi = \lambda\zeta$. One possibility would be to identify λ from the coefficients on the lags of the shadow rate and observed rate by imposing the (overidentifying) restriction that lags of Y_{2t} and Y_{2t}^* appear in the model only via the linear combination $\lambda Y_{2t}^* + (1-\lambda)$ Y_{2t}, which 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 λ. We then need an additional restriction to identify ξ so that we can draw inference on ζ. This can be done, for instance, by using external instruments, as in Gertler and Karadi (2015). Those will identify $\overline{\beta}$ directly, and, using (37), we can obtain ξ.

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 two conditions: the stability of the transmission mechanism across regimes, and the inefficacy 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 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 indicates that the ZLB is empirically relevant and unconventional policy is only partially effective.

A Proofs

A.1 Derivation of equation (15)

For any given β, the orthogonality of the errors implies that

$$\gamma = \frac{\omega_{12} - \omega_{22}\beta}{\omega_{11} - \omega_{12}\beta}, \tag{39}$$

see the proof of Proposition 3. Substituting this into (9) we have

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

So, we look at the shape of the function $g(\cdot)$.

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 $\widetilde{\beta} \neq 0$, the sign of β is the same as that of $\widetilde{\beta}$ 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 $\lim_{\beta\to\pm\infty}(\beta) = 0$. 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

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

Now, because $0 < \left(\omega_{11}\omega_{22} - \omega_{12}^2\right) < \omega_{11}\omega_{22}$ implies $\sqrt{\omega_{11}\omega_{22}\left(\omega_{11}\omega_{22} - \omega_{12}^2\right)} < \omega_{11}\omega_{22}$, 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 $\widetilde{\beta} < 0$ cannot arise from $\beta > 0$ when $\omega_{12} < 0$. In other words, observing $\widetilde{\beta} < 0$ 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 $\beta \to -\infty$, 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 $\beta_0 = \omega_{11}/\omega_{12}$, 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 $\left|\widetilde{\beta}\right| \leq |\beta|$. In other words, $\widetilde{\beta}$ is attenuated relative to the true β.

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

A.2 Proof of Proposition 1

First, a necessary condition for coherency is that the matrices $\overline{A}$ and A^* in (19) are nonsingular (otherwise, there would be fewer unknowns than equations in each or either of the two regimes and there would be no solution). Without loss of generality, we can assume that A_{11}, A_{22}^* and $A_{22}^* + A_{22}$

are nonsingular, so that the first $k-1$ equations can be solved for Y_{1t} and the last equation can be solved for Y_{2t}^* in each regime (this can always be achieved by reordering the variables in Y_t).

We need to eliminate Y_{2t}^* to obtain a system of equations in Y_{1t} and Y_{2t} alone. Let $W_t = BX_t + B^* X_t^* + \varepsilon_t$ denote the right-hand side (RHS) of (18) for compactness and solve (18) for Y_{2t}^* as a function of Y_{2t} and $W_t = (W_{1t}', W_{2t})'$, partitioned conformably with $Y_t = (Y_{1t}', Y_{2t})'$ to get

$$Y_{2t}^* = A_{22}^{*-1} \left(W_{2t} - A_{22} Y_{2t} - A_{21} Y_{1t} \right).$$

Substitute into the equation for Y_{1t} to get

$$\begin{aligned} Y_{1t} &= A_{11}^{-1} \left(-A_{12} Y_{2t} - A_{12}^* Y_{2t}^* + W_{1t} \right) \\ &= -A_{11}^{-1} A_{12} Y_{2t} - A_{11}^{-1} A_{12}^* A_{22}^{*-1} \left(W_{2t} - A_{22} Y_{2t} - A_{21} Y_{1t} \right) + A_{11}^{-1} W_{1t} \\ &= A_{11}^{-1} \left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12} \right) Y_{2t} + A_{11}^{-1} A_{12}^* A_{22}^{*-1} A_{21} Y_{1t} + A_{11}^{-1} \left(W_{1t} - A_{12}^* A_{22}^{*-1} W_{2t} \right), \end{aligned}$$

or

$$\left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right) Y_{1t} = \left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12} \right) Y_{2t} + \left(W_{1t} - A_{12}^* A_{22}^{*-1} W_{2t} \right).$$

Now, since $\det A^* = \det A_{22}^* \det \left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right) \neq 0$ (Lütkepohl, 1996, p. 50 (6)), it follows that $\left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right)$ is invertible, so

$$\begin{aligned} Y_{1t} &= \left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right)^{-1} \left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12} \right) Y_{2t} + x_{1t}, \\ x_{1t} &:= \left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right)^{-1} \left(W_{1t} - A_{12}^* A_{22}^{*-1} W_{2t} \right). \end{aligned} \quad (41)$$

Finally, the equation for Y_{2t} is

$$Y_{2t} = \max \left\{ A_{22}^{*-1} \left(W_{2t} - A_{22} Y_{2t} - A_{21} Y_{1t} \right), b \right\}.$$

Substituting for Y_{1t} using (41), we have

$$\begin{aligned} Y_{2t} &= \max \left\{ A_{22}^{*-1} \left(W_{2t} - A_{21} x_{1t} - A_{22} Y_{2t} - A_{21} \left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right)^{-1} \left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12} \right) Y_{2t} \right), b \right\} \\ &= \max \left\{ \left(x_{2t} - A_{22}^{*-1} \left(A_{22} + A_{21} \left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right)^{-1} \left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12} \right) \right) Y_{2t} \right), b \right\} \end{aligned}$$

with $x_{2t} := A_{22}^{*-1} \left(W_{2t} - A_{21} x_{1t} \right)$. Coherency and completeness hold if the above equation has a unique solution for all possible x_{2t}. If x_{2t} is arbitrary, which is the case when ε_t is supported on $\Re^k$, the necessary and sufficient coherency and completeness condition is given by

$$-A_{22}^{*-1} \left(A_{22} + A_{21} \left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21} \right)^{-1} \left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12} \right) \right) < 1. \quad (42)$$

From (Lütkepohl, 1996, p. 29 (2)), we have

$$\left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21}\right)^{-1} = A_{11}^{-1} + A_{11}^{-1} A_{12}^* \left(A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*\right)^{-1} A_{21} A_{11}^{-1}.$$

Hence,

$$\begin{aligned}
&A_{21}\left(A_{11} - A_{12}^* A_{22}^{*-1} A_{21}\right)^{-1}\left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12}\right) \\
&= A_{21} A_{11}^{-1}\left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12}\right) + A_{21} A_{11}^{-1} A_{12}^*\left(A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*\right)^{-1} \\
&\times A_{21} A_{11}^{-1}\left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12}\right) \\
&= \left(1 + \frac{A_{21} A_{11}^{-1} A_{12}^*}{A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*}\right) A_{21} A_{11}^{-1}\left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12}\right) \\
&= \frac{A_{22}^*}{A_{22}^* - A_{21} A_{11}^{-1} A_{12}^*} A_{21} A_{11}^{-1}\left(A_{12}^* A_{22}^{*-1} A_{22} - A_{12}\right)
\end{aligned}$$

Substituting back into (42) yields

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

Re-arranging yields (20).

A.3 Proof of Proposition 2

Define $\overline{A}_{i2} := A_{i2}^* + A_{i2}$, $i = 1, 2$ as the right blocks of $\overline{A}$ 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

$$Y_t^* = \begin{cases} C X_t + C^* X_t^* + u_t, & \text{if } D_t = 0 \\ \widetilde{C} X_t + \widetilde{C}^* X_t^* + \widetilde{c} b + \widetilde{u}_t, & \text{if } D_t = 1 \end{cases} \tag{43}$$

where

$$C = \overline{A}^{-1} B, \quad C^* = \overline{A}^{-1} B^*, \quad u_t = \overline{A}^{-1} \varepsilon_t \tag{44}$$

and

$$\widetilde{C} = A^{*-1} B, \quad \widetilde{C}^* = A^{*-1} B^*, \quad \widetilde{c} = -A^{*-1}\begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix} b, \quad \widetilde{u}_t = A^{*-1} \varepsilon_t. \tag{45}$$

From the partitioned inverse formula we have

$$A^{*-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1} A_{12}^{*} \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*} \right)^{-1} A_{21} A_{11}^{-1} & -A_{11}^{-1} A_{12}^{*} \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*} \right)^{-1} \\ -\left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*} \right)^{-1} A_{21} A_{11}^{-1} & \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*} \right)^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} \left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21} \right)^{-1} & -\left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21} \right)^{-1} A_{12}^{*} A_{22}^{*-1} \\ -A_{22}^{*-1} A_{21} \left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21} \right)^{-1} & A_{22}^{*-1} + A_{22}^{*-1} A_{21} \left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21} \right)^{-1} A_{12}^{*} A_{22}^{*-1} \end{pmatrix}.$$

So,

$$\widetilde{C}_1 = \left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21} \right)^{-1} \left(B_1 - A_{12}^{*} A_{22}^{*-1} B_2 \right)$$
$$\widetilde{C}_2 = \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*} \right)^{-1} \left(B_2 - A_{21} A_{11}^{-1} B_1 \right)$$

and similarly

$$C_1 = \left(A_{11} - \overline{A}_{12} \overline{A}_{22}^{-1} A_{21} \right)^{-1} \left(B_1 - \overline{A}_{12} \overline{A}_{22}^{-1} B_2 \right)$$
$$C_2 = \left(\overline{A}_{22} - A_{21} A_{11}^{-1} \overline{A}_{12} \right)^{-1} \left(B_2 - A_{21} A_{11}^{-1} B_1 \right).$$

Solving the latter for B_1 and B_2 yields

$$B_1 = \left(A_{11} - \overline{A}_{12} \overline{A}_{22}^{-1} A_{21} \right) C_1 + \overline{A}_{12} \overline{A}_{22}^{-1} B_2$$

and

$$\left(1 - A_{21} A_{11}^{-1} \overline{A}_{12} \overline{A}_{22}^{-1} \right) B_2 = \left(1 - A_{21} A_{11}^{-1} \overline{A}_{12} \overline{A}_{22}^{-1} \right) \overline{A}_{22} C_2 + \left(1 - A_{21} A_{11}^{-1} \overline{A}_{12} \overline{A}_{22}^{-1} \right) A_{21} C_1.$$

Now, since $\det \overline{A} = \det \left(A_{11} \right) \det \left(\overline{A}_{22} - A_{21} A_{11}^{-1} \overline{A}_{12} \right) \neq 0$ (Lütkepohl, 1996, p. 50 (6)), it follows that

$$\overline{A}_{22} - A_{21} A_{11}^{-1} \overline{A}_{12} \neq 0, \tag{46}$$

so

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

Thus,

$$\widetilde{C}_1 = C_1 + \left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21} \right)^{-1} \left(\overline{A}_{12} - A_{12}^{*} A_{22}^{*-1} \overline{A}_{22} \right) C_2 = C_1 - \widetilde{\beta} C_2,$$

since

$$\left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21}\right)^{-1} \left(\overline{A}_{12} - A_{12}^{*} A_{22}^{*-1} \overline{A}_{22}\right)$$
$$= -\left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21}\right)^{-1} \left(A_{12}^{*} A_{22}^{*-1} A_{22} - A_{12}\right) = -\widetilde{\beta}.$$

Next,

$$\begin{aligned}\widetilde{C}_2 &= \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*}\right)^{-1} \left(B_2 - A_{21} A_{11}^{-1} B_1\right) \\ &= \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*}\right)^{-1} \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*} + A_{22} - A_{21} A_{11}^{-1} A_{12}\right) C_2 = \kappa C_2,\end{aligned}$$

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

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

Next,

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

Finally, $\widetilde{u}_t = A^{*-1} \overline{A} u_t$, so

$$\begin{aligned}\widetilde{u}_{1t} &= \left(A_{11} - A_{12}^{*} A_{22}^{*-1} A_{21}\right)^{-1} \begin{pmatrix} I & -A_{12}^{*} A_{22}^{*-1} \end{pmatrix} \begin{pmatrix} A_{11} u_{1t} + \overline{A}_{12} u_{2t} \\ A_{21} u_{1t} + \overline{A}_{22} u_{2t} \end{pmatrix} \\ &= u_{1t} - \widetilde{\beta} u_{2t},\end{aligned}$$

and

$$\begin{aligned}\widetilde{u}_{2t} &= \left(A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*}\right)^{-1} \begin{pmatrix} -A_{21} A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} u_{1t} + \overline{A}_{12} u_{2t} \\ A_{21} u_{1t} + \overline{A}_{22} u_{2t} \end{pmatrix} \\ &= \left(1 + \frac{A_{22} - A_{21} A_{11}^{-1} A_{12}}{A_{22}^{*} - A_{21} A_{11}^{-1} A_{12}^{*}}\right) u_{2t} = \kappa u_{2t}.\end{aligned}$$

Substituting back into (43), the reduced-form model for Y_{1t} becomes

$$\begin{aligned}Y_{1t} &= (1 - D_t)\left(C_1 X_t + C_1^{*} X_t^{*} + u_{1t}\right) \\ &\quad + D_t \left(\left(C_1 - \widetilde{\beta} C_2\right) X_t + \left(C_1^{*} - \widetilde{\beta} C_2^{*}\right) X_t^{*} + u_{1t} - \widetilde{\beta} u_{2t}\right),\end{aligned} \tag{47}$$

and for Y_{2t}^* it is

$$\begin{aligned} Y_{2t}^* &= (1 - D_t)\left(C_2 X_t + C_2^* X_t^* + u_{2t}\right) + D_t\left(\kappa C_2 X_t + \kappa C_2^* X_t^* + (1-\kappa)\, b + \kappa u_{2t}\right) \\ &= C_2 X_t + C_2^* X_t^* + u_{2t} - (1-\kappa)\, D_t\left(C_2 X_t + C_2^* X_t^* + u_{2t} - b\right). \end{aligned} \tag{48}$$

Next, define

$$\widetilde{Y}_{2t}^* := C_2 X_t + C_2^* X_t^* + u_{2t}, \tag{49}$$

and rewrite (48) as

$$\begin{aligned} Y_{2t}^* &= \widetilde{Y}_{2t}^* - (1-\kappa)\, D_t\left(\widetilde{Y}_{2t}^* - b\right) \\ &= (1 - D_t)\, \widetilde{Y}_{2t}^* + D_t\left(\kappa \widetilde{Y}_{2t}^* + (1-\kappa)\, b\right). \end{aligned} \tag{50}$$

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

$$\overline{C}_{ij} = \begin{cases} C_{ij}, & j \in \{1, q\} : X_{tj} \neq Y_{2,t-s} \text{ for all } s \in \{1, p\} \\ C_{ij} + C_{is}^*, & j \in \{1, q\} : X_{tj} = Y_{2,t-s}, \text{ for some } s \in \{1, p\}. \end{cases} \tag{51}$$

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 (51), we can rewrite (49) as

$$\widetilde{Y}_{2t}^* = \overline{C}_2 X_t + C_2^* \min\left(X_t^* - b, 0\right) + u_{2t}. \tag{52}$$

Now, observe that

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

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

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

where $\overline{C}^* = \kappa C^*$. Now, from (52) we have

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

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

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

where $\overline{X}_t^{*} := (\overline{x}_{t-1}, ..., \overline{x}_{t-p})'$, and $\overline{x}_t := \min\left(\overline{Y}_{2t}^{*} - b, 0\right)$, with initial conditions $\overline{x}_{-s} = \kappa^{-1}\min\left(Y_{2,-s}^{*} - b, 0\right)$, $s = 0, ..., p-1$. It follows that $\min\left(\widetilde{X}_t^{*} - b, 0\right) = \overline{X}_t^{*}$ for all $t \geq 1$, so that $\widetilde{Y}_{2t}^{*} = \overline{Y}_{2t}^{*}$. Substituting $\overline{Y}_2^{*}$ for $\widetilde{Y}_2^{*}$ in (50), we get (25). Using the reparametrization (51) and the relationship between X_t^{*} and $\overline{X}_t^{*}$ in (47), we obtain (22).

Finally, from eq. (48), it follows that the event $Y_{2t}^{*} < b$ is equivalent to

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

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

$$u_{2t} < b - C_2 X_t - C_2^{*} X_t^{*}. \tag{53}$$

Using the definition (24), and (51), the inequality (53) can be written as $\overline{Y}_{2t}^{*} < b$, which establishes (23).

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

A.4 Proof of Proposition 3

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

$$u_{1t} = \left(A_{11} - \overline{A}_{12}\overline{A}_{22}^{-1}A_{21}\right)^{-1}\left(\varepsilon_{1t} - \overline{A}_{12}\overline{A}_{22}^{-1}\varepsilon_{2t}\right) \tag{54}$$

$$u_{2t} = \left(\overline{A}_{22} - A_{21}A_{11}^{-1}\overline{A}_{12}\right)^{-1}\left(\varepsilon_{2t} - A_{21}A_{11}^{-1}\varepsilon_{1t}\right). \tag{55}$$

Using the definitions

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

we can rewrite (54)-(55) as (30)-(31).

Note that

$$\bar{\varepsilon}_{1t} = A_{11}^{-1}\left(A_{11}u_{1t} + \overline{A}_{12}u_{2t}\right) = u_{1t} - \bar{\beta}u_{2t},$$
$$\bar{\varepsilon}_{2t} = \overline{A}_{22}^{-1}\left(A_{21}u_{1t} + \overline{A}_{22}u_{2t}\right) = -\overline{\gamma}u_{1t} + u_{2t},$$

so,

$$var\left(\bar{\varepsilon}_{1t}\right) = \left(I_{k-1}, -\bar{\beta}\right)\Omega\left(I_{k-1}, -\bar{\beta}\right)',$$
$$var\left(\bar{\varepsilon}_{2t}\right) = \left(-\bar{\gamma}, 1\right)\Omega\left(-\bar{\gamma}, 1\right)',$$

and

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

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

$$\overline{\gamma}' = \left(\Omega_{11} - \overline{\beta}\Omega_{12}'\right)^{-1}\left(\Omega_{12} - \overline{\beta}\Omega_{22}\right).$$

B Computational details

B.1 Derivation of the 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

$$s_t = \begin{pmatrix}\mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1}\end{pmatrix}, \quad \underset{(k+1)\times 1}{\mathbf{y}_t} = \begin{pmatrix}Y_t \\ \overline{Y}_{2t}^*\end{pmatrix},$$

and write the state transition equation as

$$s_t = F\left(s_{t-1}, u_t; \psi\right) = \begin{pmatrix}F_1\left(s_{t-1}, u_t; \psi\right) \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+1}\end{pmatrix}, \tag{56}$$

and

$$F_1\left(s_{t-1},u_t;\psi\right)=\begin{pmatrix}\overline{C}_1X_t+\overline{C}_1^*\overline{X}_t^*+u_{1t}-\widetilde{\beta}D_t\left(\overline{C}_2X_t+\overline{C}_2^*\overline{X}_t^*+u_{2t}-b\right)\\ \max\left(b,\overline{C}_2X_t+\overline{C}_2^*\overline{X}_t^*+u_{2t}\right)\\ \overline{C}_2X_t+\overline{C}_2^*\overline{X}_t^*+u_{2t}\end{pmatrix},$$

and the observation equation as

$$Y_t=\begin{pmatrix}I_k & 0_{k\times 1+(p-1)(k+1)}\end{pmatrix}s_t. \tag{57}$$

Next, I will derive the predictive density and mass functions conditional on the past state variables that make up the prediction error decomposition of the likelihood. The contribution of the uncensored observations is straightforward. With Gaussian errors, the joint predictive density of Y_t corresponding to the observations with $D_t=0$ can be written as:

$$f_0\left(Y_t|s_{t-1},\psi\right):=|\Omega|^{-1/2}\exp\left\{-\frac{1}{2}tr\left(\left(Y_t-\overline{C}X_t-\overline{C}^*\overline{X}_t^*\right)\left(Y_t-\overline{C}X_t-\overline{C}^*\overline{X}_t^*\right)'\Omega^{-1}\right)\right\}. \tag{58}$$

It remains to determine the contribution of the observations on the boundary, $D_t=1$. Specifically, because $D_t=1$ if and only if $u_{2t}<b-\overline{C}_2X_t-\overline{C}_2^*\overline{X}_t^*$, we need to find

$$\int_{-\infty}^{b-\overline{C}_2X_t-\overline{C}_2^*\overline{X}_t^*}f_{Y_1,u_2}\left(Y_{1t},u_{2t}|s_{t-1},\psi\right)du_{2t},$$

where f_{Y_1,u_2} is the joint density of Y_{1t} and u_{2t}, conditional on s_{t-1}. This can also be expressed as the marginal density of Y_{1t} for observations with $D_t=1$, denoted $f_1\left(Y_{1t}|s_{t-1},\psi\right)$, times $\Pr(u_{2t}<b-\overline{C}_2X_t-\overline{C}_2^*\overline{X}_t^*|Y_{1t},s_{t-1},\psi)$.

Now, at $D_t=1$, $Y_{1t}=$ (predetermined variables)$+u_{1t}-\widetilde{\beta}u_{2t}$, so $f_1\left(Y_{1t}|s_{t-1},\psi\right)$ is Gaussian, and can be written as

$$f_1\left(Y_{1t}|s_{t-1},\psi\right):=|\Xi_1|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(Y_{1t}-\mu_{1t}\right)'\Xi_1^{-1}\left(Y_{1t}-\mu_{1t}\right)\right] \tag{59}$$

$$\mu_{1t}:=\widetilde{\beta}b+\left(\overline{C}_1-\widetilde{\beta}\overline{C}_2\right)X_t+\left(\overline{C}_1^*-\widetilde{\beta}\overline{C}_2^*\right)\overline{X}_t^* \tag{60}$$

$$\Xi_1:=\Omega_{1.2}+\widetilde{\delta}\widetilde{\delta}'\tau^2=\begin{pmatrix}I_{k-1} & -\widetilde{\beta}\end{pmatrix}\Omega\begin{pmatrix}I_{k-1}\\ -\widetilde{\beta}'\end{pmatrix},\quad \widetilde{\delta}=\Omega_{12}\omega_{22}^{-1}-\widetilde{\beta}, \tag{61}$$

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

$$u_{2t}|Y_{1t},s_{t-1}\sim N\left(\mu_{2t},\tau_2^2\right),\text{ with} \tag{62}$$

$$\mu_{2t}:=\tau^2\widetilde{\delta}'\Xi_1^{-1}\left(Y_{1t}-\mu_{1t}\right),\quad \tau_2=\tau\sqrt{\left(1-\tau^2\widetilde{\delta}'\Xi_1^{-1}\widetilde{\delta}\right)}. \tag{63}$$

Hence,

$$\Pr\left(D_t = 1 | Y_{1t}, s_{t-1}, \psi\right) = \Phi\left(\frac{b - \overline{C}_2 X_t - \overline{C}_2^* \overline{X}_t^* - \mu_{2t}}{\tau_2}\right). \tag{64}$$

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

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

where $f_0(Y_t|s_{t-1}, \theta)$ and $f_1(Y_{1t}|s_{t-1}, \theta)$ are given by (58) and (59), resp., with $\overline{C}^* = 0$.[12]

The likelihood for the unrestricted CKSVAR ($\overline{C}^* \neq 0$) is not available analytically, but it can be computed approximately by simulation (particle filtering). I will describe 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 advantage of this algorithm is that the resulting likelihood approximation is continuous and smooth, so it can be readily maximized using derivative-based methods, and asymptotic standard errors can be computed numerically by the delta method. The disadvantage is that it suffers from sample degeneracy of the particles, and may give an inaccurate approximation to the true likelihood function. This can be gauged from the effective sample size (ESS) of the filter as explained below.

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). The disadvantage of any resampling algorithm is that the resulting likelihood approximation is discontinuous, and therefore this method is not amenable to derivative-based optimization or numerical computation of standard errors. Thus, I resort to simulated annealing for likelihood maximization, and use the likelihood ratio statistic for testing. This takes much longer to execute, but it is potentially more accurate than SIS.

Both algorithms require sampling from the predictive density of $\overline{Y}_{2t}^*$ conditional on $Y_{1t}, D_t = 1$ and s_{t-1}. From (24) and (62), we see that this is a truncated normal with original mean $\mu_{2t}^* = \overline{C}_2 X_t + \overline{C}_2^* \overline{X}_t^* + \mu_{2t}$ and standard deviation τ_2, where μ_{2t}, τ_2 are given in (63), i.e.,

$$f_2\left(Y_{2t}^* | Y_{1t}, D_t = 1, s_{t-1}, \psi\right) = TN\left(\mu_{2t}^*, \tau_2, \overline{Y}_{2t}^* < b\right) \tag{65}$$

Draws from this truncated distribution can be obtained using, for instance, the procedure in Lee (1999).

[12] Note that the contribution to the likelihood of the observations on the boundary is different from $\Phi\left(\frac{b - C_2 X_t}{\tau}\right)$, which would be the corresponding term in the marginal Tobit likelihood for Y_{2t}.

Let $\xi_t^{(j)} \sim U[0,1]$ be *i.i.d.* uniform random draws, $j = 1, ..., M$. Then, a draw from $\overline{Y}_{2t}^{*}|Y_{1t}, s_{t-1}, \overline{Y}_{2t}^{*} < b$ is given by

$$\overline{Y}_{2t}^{*(j)} = \mu_{2t}^{*} + \tau_2 \Phi^{-1}\left[\xi_t^{(j)} \Phi\left(\frac{b - \mu_{2t}^{*}}{\tau_2}\right)\right]. \tag{66}$$

Algorithm 1 (SIS) *Sequential Importance Sampler*

1. Initialization. *For* $j = 1 : M$, *set* $W_0^j = 1$ *and* $s_0^j = \left(\mathbf{y}_0^j, \ldots, \mathbf{y}_{-p+1}^j\right)$, *with* $\mathbf{y}_{-s}^j = \left(Y_0', Y_{2,0}\right)'$, *for* $s = 0, ..., p-1$. *(in other words, initialize* $\overline{Y}_{2,-s}^{*}$ *at the observed values of* $Y_{2,-s}$*).*

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

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

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

   where f_0, f_1, *and* $\Pr\left(D_t = 1|Y_{1t}, s_{t-1}; \psi\right)$ *are given by (58), (59), and (64), 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* $p\left(s_t|s_{t-1}^j, Y_t\right)$. *That is,* $s_t^j = \left(\mathbf{y}_t^j, \mathbf{y}_{t-1}^j, \ldots, \mathbf{y}_{t-p}^j\right)$ *where* $\mathbf{y}_t^j = \left(Y_t', \overline{Y}_{2t}^{*(j)}\right)$ *and* $\overline{Y}_{2t}^{*(j)}$ *is a draw from* $f_2\left(Y_{2t}^{*}|Y_{1t}, D_t = 1, s_{t-1}^j, \psi\right)$ *using (66).*

   (c) Update the weights:

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

3. *Likelihood approximation*

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

If the draws $\xi_t^{(j)}$ are kept fixed across different values of ψ, 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). As mentioned before, 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)

$$ESS_t = \frac{M}{\frac{1}{M}\sum_{j=1}^{M}\left(W_t^j\right)^2}. \tag{67}$$

Next, I turn to the FAPF algorithm.

Algorithm 2 (FAPF) *Fully Adapted Particle Filter*

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

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

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

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

   where f_0, f_1, *and* $\Pr(D_t = 1|Y_{1t}, s_{t-1}; \psi)$ *are given by (58), (59), and (64), 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* $\left\{j, \pi_{t-1|t}^j\right\}$. *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,* $\tilde{s}_{t-1}^j = \left(\tilde{\mathbf{y}}_{t-1}^j, \ldots, \tilde{\mathbf{y}}_{t-p}^j\right)$, $\tilde{\mathbf{y}}_{t-s}^j = \left(Y_{t-1}', \overline{Y}_{2,t-s}^{*(k_j)}\right)$, $s = 1, ..., p$.*)*

   (c) For $j = 1 : M$, *sample* s_t^j *randomly from* $p\left(s_t|\tilde{s}_{t-1}^j, Y_t\right)$. *That is,* $s_t^j = \left(\mathbf{y}_t^j, \tilde{\mathbf{y}}_{t-1}^j, \ldots, \tilde{\mathbf{y}}_{t-p}^j\right)$ *where* $\mathbf{y}_t^j = \left(Y_t', \overline{Y}_{2t}^{*(j)}\right)$ *and* $\overline{Y}_{2t}^{*(j)}$ *is a draw from* $f_2\left(Y_{2t}^*|Y_{1t}, D_t = 1, \tilde{s}_{t-1}^j, \psi\right)$ *using (66).*

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 $h\left(s_t^j\right)$ produce the requisite average or quantiles of $h(s_t)$ conditional on $Y_1, \ldots, Y_t$ (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 $h\left(s_t^j\right)$ can be obtained using weights W_T.

B.2 Other estimators

B.2.1 Method of moments

A method of moments estimator for ψ can be constructed using equation (27), together with the following equations

$$E\left(Y_{2t}|X_t, \overline{X}_t^*\right) = \overline{C}_2 X_t + \overline{C}_2^* \overline{X}_t^* + \Phi\left(a_t\right)\left(b - \overline{C}_2 X_t - \overline{C}_2^* \overline{X}_t^*\right) + \tau\phi\left(a_t\right), \tag{68}$$

$$E\left(D_t|X_t\right) = \Phi\left(a_t\right), \quad a_t = \frac{b - \overline{C}_2 X_t - \overline{C}_2^* \overline{X}_t^*}{\tau} \tag{69}$$

and an expression for $var\left(Y_{1t}|X_t, \overline{X}_t^*, Y_{2t}\right)$. When $C^* \neq 0$, we need to integrate out $\overline{X}_t^*$ from the above expressions. Then, the regressors become $\mathbf{X}_t := \left(X_t', \overline{X}_{t|t}^{*\prime}, Z_{1t|t}, Z_{2t|t}\right)'$, where $h_{t|t} := E\left(h\left(\overline{X}_t^*\right)|Y_{2t}, Y_{t-1}, ..., Y_1\right)$ for any function $h\left(\cdot\right)$ whose expectation exists. Structural estimates can be obtained by the same transformations as for ML.

B.2.2 Sequential estimation of KSVAR

The KSVAR (with $C^* = 0$ and $\lambda = 0$, so $\widetilde{\beta} = \beta$) can be estimated easily using standard estimation routines that are available in most econometrics software. Specifically, the following sequential estimator involves the steps Tobit-OLS-Tobit:

1. Estimate the reduced-form parameters ψ_2 by ML from the Tobit regression of Y_{2t} given by equations (23) and (24) with $\overline{C}_2^* = 0$. Denote them $\widehat{\psi}_2$.

2. Construct the auxiliary variables $\widehat{Z}_{1t} = Z_{1t}\left(\widehat{\psi}_2\right)$, $\widehat{Z}_{1t} = Z_{2t}\left(\widehat{\psi}_2\right)$ defined in (28)-(29) with $\overline{C}_2^* = 0$, and estimate $C_1 = \overline{C}_1, \beta = \widetilde{\beta}$, and δ from (27) (with $C_1^* = 0$) by OLS. Denote the estimate of β by $\widehat{\beta}$.

3. Construct $\widehat{Z}_t = Y_{1t} - \widehat{\beta} Y_{2t}$ and estimate the Tobit regression (36) (with $\tilde{B}_2^* = 0$) by ML. This yields $\widehat{\widetilde{\gamma}}$, $\widehat{\widetilde{B}}_2$, whereupon you get estimates of the structural parameters $\widehat{\gamma} = \widehat{\widetilde{\gamma}}\left(I_{k-1} + \widehat{\beta}\widehat{\widetilde{\gamma}}\right)^{-1}$, and $\widehat{B}_2 = \left(1 - \widehat{\gamma}\widehat{\beta}\right)\widehat{\widetilde{B}}_2$.

B.3 Computation of the identified set

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

$$\widetilde{\beta} = (1 - \xi)\left(I - \xi\overline{\beta}\left(\Omega_{12}' - \Omega_{22}\overline{\beta}'\right)\left(\Omega_{11} - \Omega_{12}\overline{\beta}'\right)^{-1}\right)^{-1}\overline{\beta}. \tag{70}$$

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 $\overline{\beta}$ can then be obtained numerically by solving (70) for $\overline{\beta}$ as a function of the reduced-form parameters and ξ for each value of ξ, and gathering all the solutions in the set. It is shown in the supplementary appendix that (70) can be written as

$$\widetilde{\beta} - \tilde{A}\overline{\beta} + \overline{\beta}\overline{\beta}'\tilde{b} = 0, \tag{71}$$

where

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

Defining $z := \tilde{b}'\overline{\beta}$, it is further shown in the supplementary appendix that premultiplying equation (71) by $\tilde{b}'$ yields the equation

$$\begin{aligned} 0 = {} & \tilde{b}'\widetilde{\beta}\det\left(C_0\left(z\right)\right) + \tilde{b}'\tilde{A}\tilde{b}\left(\tilde{b}'\tilde{b}\right)^{-1} z\det\left(C_0\left(z\right)\right) \\ & - \tilde{b}'\tilde{A}\tilde{b}_{\perp}C_0\left(z\right)^{adj} c_1\left(z\right) + \det\left(C_0\left(z\right)\right)z^2, \end{aligned} \tag{72}$$

where

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

$\tilde{b}_{\perp}$ 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 $\overline{\beta}$ are given by

$$\overline{\beta}_i = \tilde{b}\left(\tilde{b}'\tilde{b}\right)^{-1} z_i + \tilde{b}_{\perp}C_0\left(z_i\right)^{-1}c_1\left(z_i\right). \tag{73}$$

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

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 (72).*

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

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

(a) *derive* $\overline{\beta}_{i,r}$ *from (73),* $\overline{\gamma}_{i,r} = \left(\Omega_{12}' - \Omega_{22}\overline{\beta}_{i,r}'\right)\left(\Omega_{11} - \Omega_{12}\overline{\beta}_{i,r}'\right)^{-1}$,
$\overline{A}_{22,i,r}^{-1} = \sqrt{\left(-\overline{\gamma}_{i,r}, 1\right)\Omega\left(-\overline{\gamma}_{i,r}, 1\right)'}$, *and* $\Xi_{1,i.r} = \left(I_{k-1}, -\overline{\beta}_{i,r}\right)\Omega\left(I_{k-1}, -\overline{\beta}_{i,r}\right)'$;

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

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

ii. *for any scalar* ς, *set*

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

and compute $Y_{t,i,r}^{j}\left(\varsigma\right)$ *using (22)-(23) with* $u_{t,i,r}^{j}\left(\varsigma\right)$ *in place of* u_t, *and iterate forward to obtain* $Y_{t+h,i,r}^{j}\left(\varsigma\right)$ *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* $\overline{\varepsilon}_{2t}$, *or* $\varsigma = \overline{A}_{22,i,r}^{-1}$ *for a one-standard deviation impulse.*

(c) *compute*

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

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

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. The parameters are set as follows. 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 $B_{1,1} = \left(\rho I_2, 0\right)$, $B_{1,1}^{*} = 0_{2\times 1}$, $B_{2,1} = \left(0_{1\times 2}, B_{22,1}\right)$, with $\rho = 0.5$. Finally, in each of the three DGPs is determined as follows. DGP1: $B_{22,1} = B_{2,1}^{*} = 0$ (both KSVAR and CSVAR, since lags of $Y_{2,t}$ and $Y_{2,t}^{*}$ all have zero coefficients); DGP2: $B_{22,1} = \rho$, $B_{2,1}^{*} = 0$ (KSVAR but not CSVAR); DGP3: $B_{22,1} = 0$, $B_{2,1}^{*} = \rho$ (CSVAR but not KSVAR). 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}

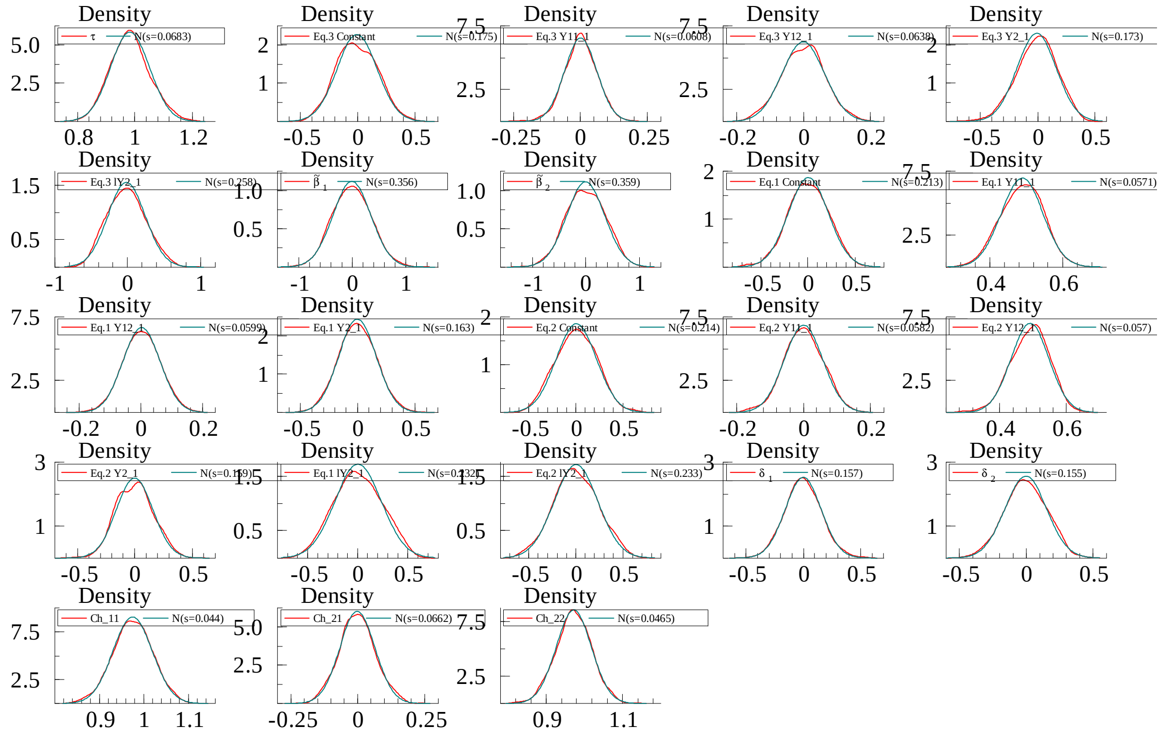

Figure 5: Sampling distribution of reduced-form coefficients of CKSVAR(1) in DGP1. $T = 250$, 1000 Monte Carlo replications.

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.

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 2, 3 and 4 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 RMSE declines at rate $\sqrt{T}$ in accordance with asymptotic theory. It is noteworthy that the estimators of $\widetilde{\beta}$ 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 for of the restrictions 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 χ^2_3 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 $\widetilde{\beta} = 0$), and the LR statistic is asymptotically distributed

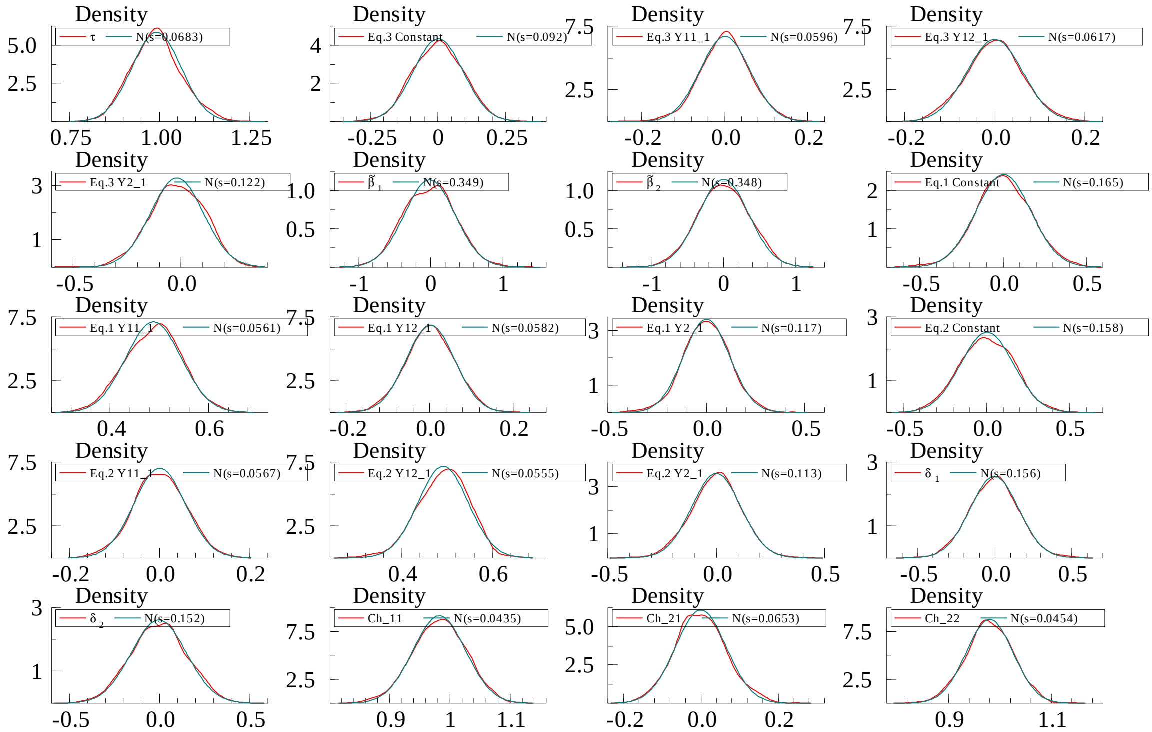

Figure 6: Sampling distribution of reduced-form coefficients of KSVAR(1) in DGP1. $T = 250$, 1000 Monte Carlo replications.

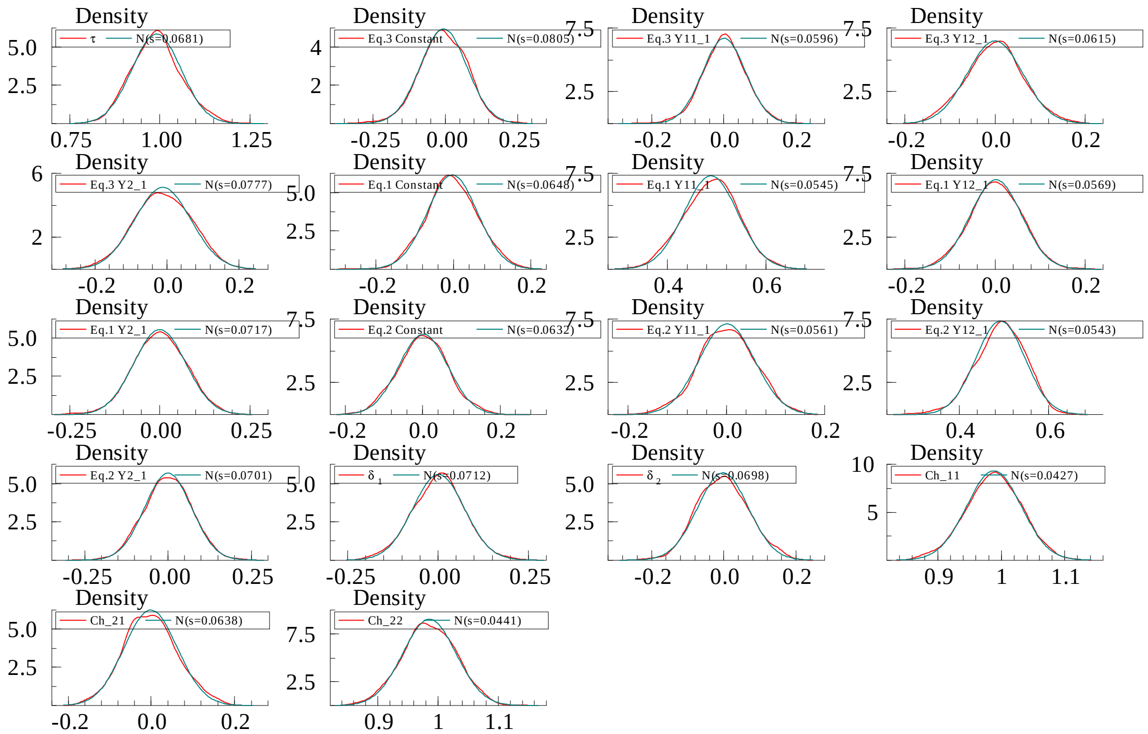

Figure 7: Sampling distribution of reduced-form coefficients of CSVAR(1) in DGP1. $T = 250$, 1000 Monte Carlo replications.

ML-CKSVAR	true	mean	bias	sd	RMSE
τ	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
δ_1	0.000	-0.001	-0.001	0.157	0.157
δ_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 2: Moments of sampling distribution of ML estimator of the parameters CKSVAR(1), under DGP1. $T = 250$, 1000 MC replications.

ML-KSVAR	true	mean	bias	sd	RMSE
τ	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
δ_1	0.000	-0.003	-0.003	0.156	0.156
δ_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 3: Moments of sampling distribution of ML estimator of the parameters KSVAR(1), under DGP1. $T = 250$, 1000 MC replications.

ML-CSVAR	true	mean	bias	sd	RMSE
τ	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
δ_1	0.000	0.002	0.002	0.071	0.071
δ_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 4: Moments of sampling distribution of ML estimator of the parameters CSVAR(1), under DGP1. $T = 250$, 1000 MC replications.

		H_0 : KSVAR, H_1 : CKSVAR			H_0 : CSVAR, H_1 : CKSVAR		
	Sign. Level	10%	5%	1%	10%	5%	1%
DGP1	asymptotic	0.232	0.140	0.041	0.236	0.142	0.038
	bootstrap	0.105	0.048	0.014	0.114	0.048	0.006
DGP2	asymptotic	0.187	0.115	0.036	**0.250**	**0.149**	**0.049**
	bootstrap	0.11	0.053	0.014	**0.142**	**0.091**	**0.027**
DGP3	asymptotic	**0.387**	**0.297**	**0.117**	0.237	0.135	0.035
	bootstrap	**0.283**	**0.161**	**0.046**	0.108	0.047	0.006

Table 5: 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 χ^2_3 and χ^2_5 critical values for KSVAR and CSVAR resp. The bootstrap rej. frequencies were computed uisng the warp-speed method of Giacomini et al. (2013). Bold numbers indicate that the rejecction frequencies were computed under H_1 (power).

as χ^2_5. Table 5 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 is obtained using Normal errors draws and the estimated reduced-form parameters to generate the bootstrap samples. 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 sample gets large, but this level of overrejetion is clearly unsatisfactory at $T = 250$ which is a typical sample size one encounters with macroeconomic data. 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 5 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, focusing on the power of the size-correct bootstrap tests, we notice the LR test is twice as likely to detect the deviation from KSVAR than the deviation from CSVAR.

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, and F. Schorfheide (2016). Macroeconomic dynamics near the zlb: A tale of two countries. Technical report, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

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.

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.

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.

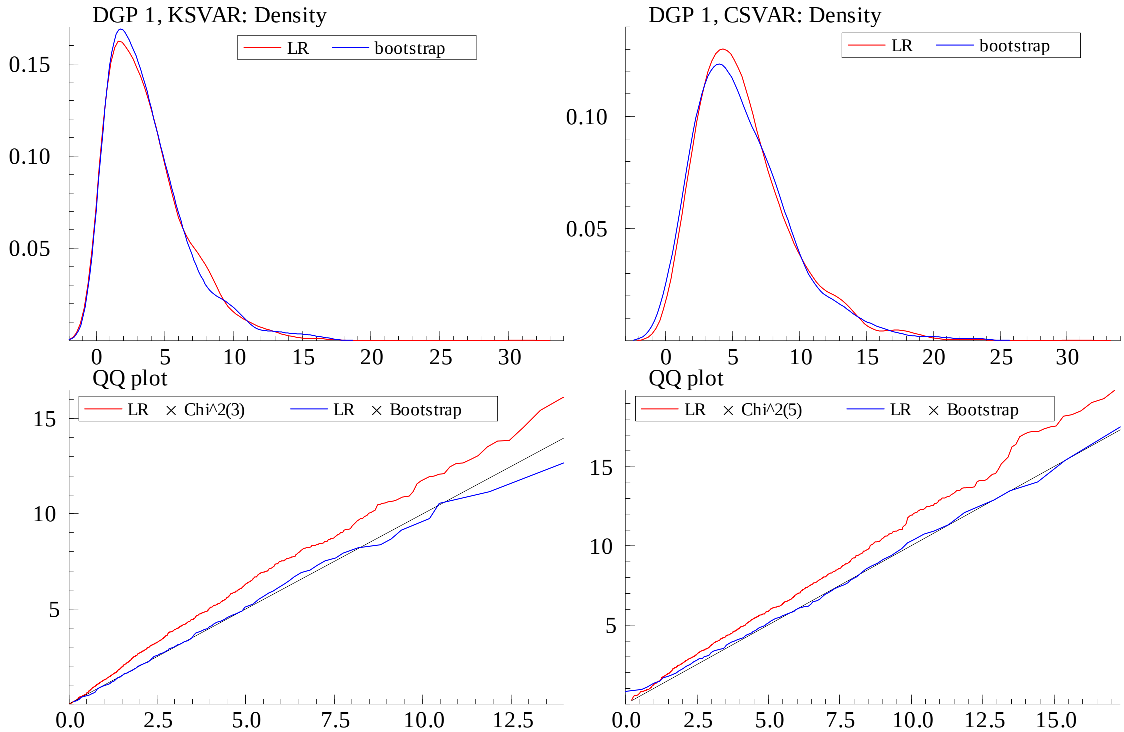

Figure 8: Sampling densities and QQ plots of LR statistics of KSVAR against CKSVAR (left) or CSVAR against CKSVAR (right) in red. Bootstrap densities in blue. Computed for $T = 250$ using 1000 Monte Carlo replications.

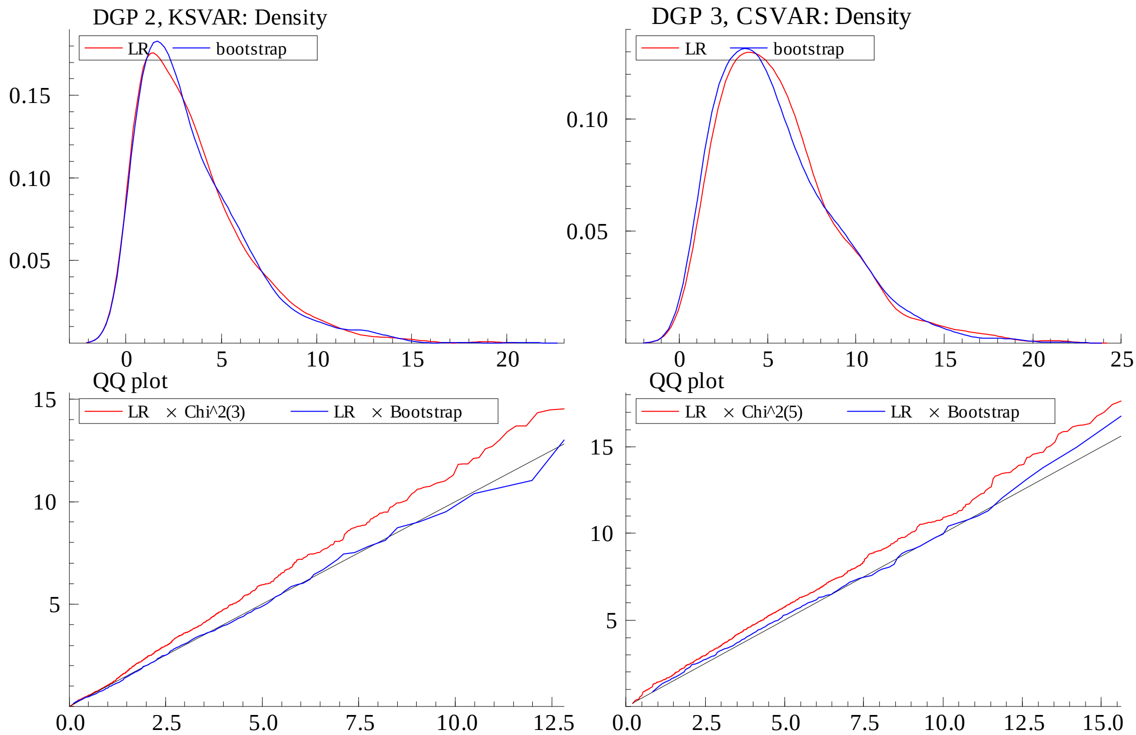

Figure 9: Sampling densities and QQ plots of LR statistics of KSVAR against CKSVAR (left) or CSVAR against CKSVAR (right) in red. Bootstrap densities in blue. Computed for $T = 250$ using 1000 Monte Carlo replications.

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. Galí, and L. Gambetti (2018). On the empirical (ir) relevance of the zero lower bound constraint. Technical report, Barcelona Graduate School of Economics.

Eggertsson, G. B. et al. (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-Ramirez (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: Journal of the Econometric Society*, 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 (2013). A regime-switching svar analysis of quantitative easing. Working Paper CARF-F-322, Center for Advanced Research in Finance, Faculty of Economics, The University of Tokyo.

Hayashi, F. and J. Koeda (2014). Exiting from qe. Working Paper 19938, National Bureau of Economic Research.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckman, J. J. et al. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

Herbst, E. P. and F. Schorfheide (2015). *Bayesian estimation of DSGE models.* Princeton and Oxford: Princeton University Press.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics 74*(1), 119–147.

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.

Lucas, R. E. J. (1976). Econometric policy evaluation: a critique. In K. Brunner and A. Meltzer (Eds.), *The Phillips Curve and Labor Markets.*, Carnegie-Rochester Conference Series on Public Policy. Amsterdam: North-Holland.

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.

Mertens, K. and M. O. Ravn (2014). Fiscal policy in an expectations driven liquidity trap. *Review of Economic Studies*. forthcoming.

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.

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. (2018). The federal reserve is not very constrained by the lower bound on nominal interest rates. Technical report, National Bureau of Economic Research.

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.