The Statistical Implications of Common Identifying Restrictions for DSGE Models*

Stephen D. Morris†
Department of Economics
UC San Diego

First Draft: September 16, 2013
This Draft: February 14, 2014
`Link for Most Recent Draft`

Abstract

I reveal identification failures in a well-known dynamic stochastic general equilibrium (DSGE) model, and study the statistical implications of common identifying restrictions. First, I provide a fully analytical methodology for determining all observationally equivalent values of the structural parameters in any parameter space. I show that either parameter admissibility or sign restrictions may yield global identification for some parameter realizations, but not for others. Second, I derive a "plug-in" maximum likelihood estimator, which requires no numerical search. I use this tool to demonstrate that the idiosyncratic identifying restriction directly impinges on both the location and distribution of the small-sample MLE, and compute correctly sized confidence intervals.

JEL Classifications: C32, C51, E12, E52.
Keywords: DSGE, Identification, Maximum Likelihood.

*I would like to thank my advisor at UC San Diego, James D. Hamilton, for guidance and support. I also benefited from feedback from members of my dissertation committee, as well as numerous seminar participants. `Supplementary material` is available on the author's website, as is a `complete derivation` of the main model utilized in this paper and all implemented `MATLAB and C code`.

†University of California, San Diego, Department of Economics, 9500 Gilman Drive #0508, La Jolla, CA 92093-0508. `sdmorris@ucsd.edu`.

1 Introduction

DSGE models are the workhorse of modern macroeconomics. They are taught in nearly all graduate economics programs, and are a core empirical tool of monetary policymakers and academics alike.[1] Following the realization of identification failures in the classic vintage of estimated multi-equation macroeconomic models by Sims (1980), calibration was suggested for what are now known as DSGE models by Kydland and Prescott (1982).[2] Eventually, increases in computing power made DSGE likelihood computation feasible, leading to the estimation of DSGE models, in the early 2000's.[3] Since then, important policies are routinely made on the basis of estimates of DSGE parameters, which typically include the discount rate, coefficient of relative risk aversion, indices of price and wage stickiness, and other theoretical objects. Recently, however, the identifiability of DSGE parameters has been called into question.[4]

In this paper, I investigate whether it is possible to distinguish between parameter values in DSGE models on the basis of the data, what may be done when this is not so, and what the statistical implications of the corrective actions are. I conduct my study using the very well-known model utilized in An and Schorfheide (2007). Although relatively small in scale, this model includes features of the seminal frameworks of Smets and Wouters (2003), Woodford (2003), and Christiano et al. (2005). Furthermore, these baseline specifications are the starting point for recent extensions, including the introduction of nonlinearity, non-normality, and more richly developed labor and financial markets.[5] In this paper, I do not venture to derive econometric results that are widely applicable to all DSGE models. Rather, I hand-craft the discussion to the

[1] Of the many central banks that openly use DSGE models to inform monetary policy decisions are the Swedish Sveriges Riksbank, the Norwegian Norge Bank, and the US Federal Reserve. See also Christiano et al. (2010). In terms of academic diffusion, as of February 2014, the representative paper of Smets and Wouters (2003) has 2,755 scholarly citations on Google Scholar.

[2] The so-called "classic vintage" including FRB-MIT. See Rasche and Shapiro (1968).

[3] Notably, Smets and Wouters (2003) and Ireland (2004).

[4] Cochrane (2011) considers the identification of the Taylor rule, Kleibergen and Mavroeidis (2009) the Phillips curve, and Beyer and Farmer (2006) and Canova and Sala (2009) the complete systems of equations known as DSGE models. Thorough critiques of the DSGE paradigm in general have also been voiced (Chari et al. (2009)). Identification is of preliminary importance for any argument for or against empirical efficacy.

[5] Galí et al. (2011) include unemployment, Bianchi (2013) allows for regime-switching, and Doh (2011), van Binsbergen et al. (2012), and Rudebusch and Swanson (2012) consider nonlinearity and the term structure.

representative model of An and Schorfheide, and in this setting am able to make clear and concise statements about problems that besiege a much wider literature, but have previously gone underappreciated.

The main findings are three-fold. First, I confirm that the conditional identification scheme suggested by Komunjer and Ng (2011) for the An and Schorfheide model ensures local identification, but go on to prove that it does *not* imply global identification.[6] In particular, for any value of the structural parameters there is exactly one other value which yields an identical value for the likelihood function, regardless of data sample. Second, I show how this problem might feasibly be addressed using common identifying restrictions based on macroeconomic theory, including parameter admissibility and sign restrictions.[7] Third, I show that these restrictions endogenously affect both the placement and distribution of the small-sample maximum likelihood estimator.[8] Thus, I echo an important recent result of Ríos-Rull et al. (2012) that the identifying restrictions themselves characterize important features of the estimator. I add to the discussion by computing correctly sized confidence intervals in this context.

In order to discuss identification rigorously, it is necessary to indicate the formal definitions I have in mind. I discuss these before placing my contribution in the context of the wider literature.

2 Definitions

DSGE models are cross-equation and exclusion restrictions on systems of time series. Let Y be a $Tn_Y \times 1$ vector of T observations of the $n_Y \times 1$ vector of data Y_t and θ be an $n_\theta \times 1$ vector structural parameter. For each $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, define a continuous likelihood function $\ell(\theta; Y)$. Rothenberg (1971) gives the following definitions:[9]

[6] "Conditional identification" is most descriptively defined by "fixing some parameters to constants."

[7] For example, one salient admissibility constraint is that the discount factor β must be less than 1. A sign restriction is that a decrease in interest rates causes the impulse-response of output to react positively.

[8] Generally, DSGE models (an in particular, An and Schorfheide (2007)) are estimated using Bayesian estimators. Thus, the discussion of the MLE might immediately seem unimportant, or at least of secondary interest. I discuss the importance of understanding the properties of the MLE, and its implications for Bayesian estimators, subsequently.

[9] See p. 578 Definitions 1-3. I use the emphasis *global* identification in Definition 3, as does Rothenberg, beginning on page 579.

Definition 1: Two parameter points θ_0 and θ_0^* are said to be *observationally equivalent* if $\ell(\theta_0; Y) = \ell(\theta_0^*; Y)$ for all $Y \in \mathbb{R}^{Tn_Y}$.

Definition 2: A parameter point θ_0 is said to be *locally identifiable* if there exists an open neighborhood of θ_0 containing no other $\theta_0^* \in \Theta$ which is observationally equivalent.

Definition 3: A parameter point θ_0 is said to be *globally identifiable* if there is no other $\theta_0^* \in \Theta$ which is observationally equivalent.

Both local and global identifiability are negatively defined in terms of observational equivalence. Observational equivalence, in turn, does not depend on the data set utilized. Thus, both local and global identifiability are features of the model, and not a data set. In addition, whereas local identifiability is qualified by uniqueness only in an open neighborhood, global identifiability is uniqueness in the entire parameter space Θ. Therefore, global identifiability is a strictly stronger assumption than local.

A concept closely linked to parameter identifiability is reduced form representation. While this term has several possible interpretations based on context, I refer to a specific meaning. In particular, Rothenberg (1971) also presents the following definition:[10]

Definition 4: Let $\theta \in \Theta$ be vector of structural parameters, and say that there exists an n_Π-dimensional continuously differentiable vector-valued function $\Pi = g(\theta)$ mapping Θ into $\mathbb{R}^{n_\Pi}$ such that $\ell(\theta; Y) = \ell^*(\Pi; Y)$ for all $Y \in \mathbb{R}^{Tn_Y \times 1}$ and $\theta \in \Theta$. If Π is globally identified in the image of Θ under g for every $\theta \in \Theta$, Π is a *reduced form parameter.*

Reduced form parameters are useful when they are available, since they completely characterize the likelihood; $\ell(\theta; Y) = \ell^*(\Pi; Y)$. For example, if Π is a reduced form parameter, and if $g(\theta_0) = g(\theta_0^*)$, then $\ell(\theta_0; Y) = \ell^*(g(\theta_0); Y) = \ell^*(g(\theta_0^*); Y) = \ell(\theta_0^*; Y)$ for all $Y \in \mathbb{R}^{Tn_y \times 1}$ by identity. In fact, using this observation, we have the following

[10] See Assumptions VII and VIII, pp. 584-5. While the definition of a reduced form parameter is rather technical, there are many instructive examples of reduced form parameters; the simplest is the 2×1 vector $\Pi = (\mu, \sigma)'$, where μ and σ are the mean and standard deviation of a univariate Gaussian likelihood, respectively.

immediate corollary to Definitions 1 and 4.

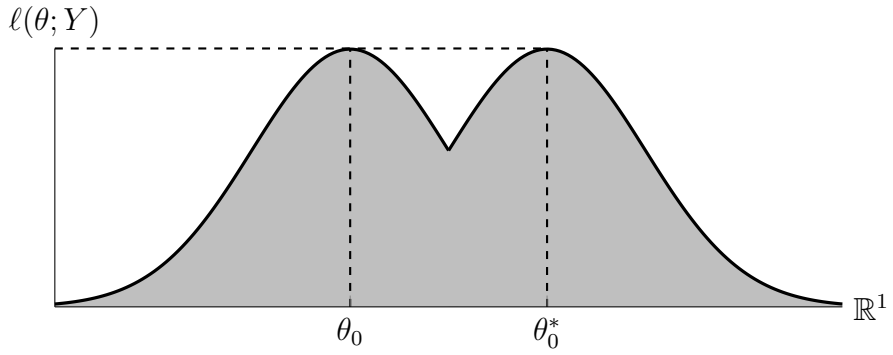
> **Corollary 1:** Let $\Pi = g(\theta)$ be a reduced form parameter. Two structural parameters θ_0 and θ_0^* are observationally equivalent if $g(\theta_0) = g(\theta_0^*)$.

A particularly useful aspect of Corollary 1 is that it can be used as an alternative to Definition 1 for the purpose of defining identification. In other words, when a given model has reduced form representation, both local and global identification depend entirely on the parametric features of $g(\theta)$. In addition, all characterizations of local and global identification in terms of reduced form representation hold in samples of variable length T, including in population.

While the contribution of Rothenberg (1971) is by now a classic, identification of parametric econometric models has an even longer history, dating back to at least Koopmans and Reiersøl (1950), Wald (1950), and Fisher (1966). What makes DSGE models special – and complicated – is the fact that it is difficult to apply the classic results. The reasons are twofold. On the one hand, DSGE solutions are not usually found analytically. Thus, for any value θ_0 it is typically not possible to directly find all θ_0^* such that $\ell(\theta_0; Y) = \ell(\theta_0^*; Y)$ exactly for all $Y \in \mathbb{R}^{Tn_Y}$. In other words, it is difficult to find the set of points which are observationally equivalent, as they are defined in Definition 1. On the other hand, the ABCD representation that DSGE models typically have is not a reduced form as defined in Definition 4 (See Fernández-Villaverde et al. (2007) and Komunjer and Ng (2011)). So, Corollary 1 may not be used.

An important preliminary point of clarification is why global identification of DSGE models matters in the first place. One reason is that observational equivalence can cause there to be multiple likelihood-maximizing parameter values in the admissible parameter space for a given sample. Say that $n_\theta = 1$, and the likelihood maximizer in a specific sample $Y \in \mathbb{R}^{Tn_Y \times 1}$ is $\theta_0 = \arg\max \ell(\theta; Y)$ where $\theta \in \mathbb{R}^1$. However, this value θ_0 has one observationally equivalent point, $\theta_0^* \neq \theta_0$ but $\theta_0^* \in \mathbb{R}_1$. Thus, $\theta_0^* = \arg\max \ell(\theta; Y)$ by definition, and the likelihood will have two equally tall peaks in the domain $\mathbb{R}^1$, as in Figure 1. This would cause the results of numerical search for a unique maximum to be misleading. Yet it is important to take note that conversely,

Figure 1: Likelihood for data sample $Y \in \mathbb{R}^{Tn_Y \times 1}$ under observational equivalence between the scalar likelihood maximizers $\theta_0 = \arg\max \ell(\theta; Y)$ and $\theta_0^* = \arg\max \ell(\theta; Y)$, but $\theta_0^* \neq \theta_0$.



two peaks of the likelihood for a given Y necessarily implies neither observational equivalence nor lack of global identification. Rather, this equality is only necessarily a property of the specific sample Y. In another sample, that phenomenon might cease.

In the next section, I utilize the above definitions to place my paper in context of the wider literature.

3 Contributions to the Literature

This paper is part of a small, but important literature which considers identification of DSGE models. Following the realization of problems by Beyer and Farmer (2006) and Canova and Sala (2009), conditions for local identification were provided by Iskrev (2010), Komunjer and Ng (2011), and Qu and Tkachenko (2012). However, local identification is merely necessary, and not sufficient for global identification; in terms of Definitions 2 and 3, there may be observationally equivalent structural parameters outside of an open neighborhood of the point of interest. In general, it is very difficult to state necessary and sufficient conditions for global identification in nonlinear models, and only the overly strong sufficient conditions of Gale and Nikaidô (1965) are typically useful. As a result, there exist no published results on global identification of DSGE models, although working papers by Fukač et al. (2007), Kocięcki and Kolasa (2013), and Qu and Tkachenko (2013) have managed to make strides in this direction.

The three aforementioned papers on global identification are the most generally use-

ful contributions on the topic thus far. Yet, there are a number of significant caveats yet to be overcome. First, all existing approaches suggest some form of searching a given parameter space for observationally equivalent points. Such a numerical search is daunting even in a small multidimensional parameter space, and infeasible in a desirably large one. Furthermore, the computational toll necessary to search for even one observationally equivalent point makes repeating the process for many points impractical. Second, the numerical algorithms necessary to solve DSGE models insert small errors in the mapping from structural parameters to moments which may lead two observationally equivalent points to seem distinct, or vice-versa. Third, and possibly most importantly, the only solutions for global identification failures that have been suggested are to fix some parameters to constants, and/or to restrict the admissible parameter space. The statistical implications of these common procedures are not well understood. The importance of uncovering global identification failures seems dubious if the methods for correcting that identification failure induce biases that are themselves unknown.

The ultimate reason for these difficulties is that DSGE models neither have analytical solutions nor reduced form representation in general. This is rightly a preliminary assumption of all of the papers of Komunjer and Ng (2011), Kocięcki and Kolasa (2013), and Qu and Tkachenko (2013), which also all use the model of An and Schorfheide (2007) to demonstrate their methodologies. In this paper, I do not offer generally applicable results which address all problems above; achieving both at once seems implausible. Rather, I simply point out that main model that has been studied in this literature, An and Schorfheide's, has both an analytical solution *and* reduced form representation. These two facts allow me to address all of the problems discussed above using simple econometric tools, and the resulting discussion is both concise and transparent.

Existing papers have only been able to evaluate global identifiability of a few representative points, subject to numerical error, in small parameter spaces, using brute-force computation. I am able to evaluate the global identifiability of hundreds of thousands of points, subject to no numerical error, in any parameter space, in a mat-

ter of minutes. Furthermore, the analytical mapping has a second use as a "plug-in" maximum likelihood estimator. Efficient estimators of the reduced form parameters are found by ordinary least squares estimators, and simply input into the mapping to yield the maximum likelihood estimators of the structural parameters. Usually, the MLE for the structural parameters is especially difficult to find for DSGE models due to numerical difficulty (See Andreasen (2010)). Since the guaranteed likelihood-maximizing values are found with certainty in a fraction of a second using the plug-in MLE, it becomes possible to compute small-sample standard errors conditional on each specific identification scheme.

To stress, these contributions should be considered complementary, rather than in substitution of, the more generally applicable econometric results authors have ventured to derive elsewhere; the results presented here are indeed merely representative of problems and consequences that may arise in more general cases. However this paper is useful in itself from a conceptual standpoint, since it clarifies problems that are difficult to uncover otherwise.

In the next section, I derive this analytical solution and reduced form representation. Then, I show how these results may be used to conduct global identification analysis.

4 Solution and Reduced Form Representation

Interest rates, output, and inflation are three of the most important aggregate variables from the perspective of monetary policy formulation, and empirical macroeconomic analysis more broadly. Consider, then, the simplest model of their dynamic relationship: An unrestricted Gaussian VAR(1) of the logged deviation from unconditional means of the nominal interest rate, r_t, detrended nominal output, y_t, and inflation, π_t.

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \phi_{rr} & \phi_{ry} & \phi_{r\pi} \\ \phi_{yr} & \phi_{yy} & \phi_{y\pi} \\ \phi_{\pi r} & \phi_{\pi y} & \phi_{\pi\pi} \end{bmatrix}}_{\Phi} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \qquad \Omega \equiv E(U_t U_t') = \begin{bmatrix} \omega_r^2 & \cdot & \cdot \\ \omega_{yr} & \omega_y^2 & \cdot \\ \omega_{\pi r} & \omega_{\pi y} & \omega_\pi^2 \end{bmatrix} \tag{1}$$

U_t is mean zero. By the Yule-Walker equations, $\Phi = \Sigma_Y(1)\Sigma_Y(0)^{-1}$ and the variance-covariance matrix is $\Omega = \Sigma_Y(0) - \Sigma_Y(1)\Sigma_Y(0)^{-1}\Sigma_Y(1)'$, where $\Sigma_Y(i) = E(Y_t Y_{t-i}')$. The unique elements of the covariance matrices, $\text{vech}(\Sigma_Y(0))$ and $\Sigma_Y(1)$, are in some sense the most primitive reduced form parameters, since the likelihood function for any vector zero-mean Gaussian process may be written as a function of its second moments. Meanwhile, the coefficient Φ, for example, is possibly more naturally interpreted as the projection of Y_t on Y_{t-1}. However, at the same time we have $\text{vech}(\Sigma_Y(0)) = \left(D_3^+(I_3 - \Phi \otimes \Phi)D_3\right)^{-1} \text{vech}(\Omega)$ and $\Sigma_Y(1) = \Phi\Sigma_Y(0)$.[11] So, the relationship between the 15 elements of the second moments $(\text{vech}(\Sigma_Y(0)), \Sigma_Y(1))$ and the 15 VAR parameters $(\Phi, \text{vech}(\Omega))$ is one-to-one; in one direction, since the second moments are functions of the VAR parameters, if the likelihood may be written as a function of the second moments, it is just as easily written as a function of the VAR parameters. In the other direction, since the VAR parameters are functions of the second moments, they are trivially globally identified in all closed and compact 15-dimensional real spaces $\Theta \subset \mathbb{R}^{15}$ in both sample and population, regardless of their realization. Thus, the VAR parameters are just as easily interpreted as reduced form parameters, and in addition, efficient estimators $\widehat{\Phi}$ and $\widehat{\Omega}$ are trivially available from OLS.

The unrestricted Gaussian VAR(1) with 15×1 reduced form parameter $\Pi = (\text{vec}(\Phi)', \text{vech}(\Omega)')'$ is ideal. But how does it compare with a standard DSGE model? The linearized equilibrium equations for the An and Schorfheide (2007) model are given by the following 6 equilibrium equations.[12] Variable and parameter names are given in Table 1.

$$z_t = \rho_z z_{t-1} + \varepsilon_{zt} \tag{2}$$

$$g_t = \rho_g g_{t-1} + \varepsilon_{gt} \tag{3}$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_\pi \pi_t + (1-\rho_r)\psi_y(y_t - g_t) + \varepsilon_{rt} \tag{4}$$

$$y_t = E_t y_{t+1} + g_t - E_t g_{t+1} - (1/\tau)(r_t - E_t \pi_{t+1} - E_t z_{t+1}) \tag{5}$$

[11] D_{n_Y} is the duplication matrix, $D_{n_Y}^+$ is its Moore-Penrose pseudoinverse, and $\otimes$ is the Kronecker product. See Abadir and Magnus (2005).

[12] A complete derivation of this model from microfoundations is a available at `https://drive.google.com/file/d/0B_TKtsickRXDU2lNSmZPbURLbDQ/edit?usp=sharing`.

	Structural Params (16)		**Endogenous** (6)		**Innovations** (3)	
1	τ	CRRA	z_t	Total factor prod.	ε_{zt}	Inn. to z_t
2	β	Discount factor	g_t	Gov spending	ε_{gt}	Inn. to g_t
3	ν	Inverse elas. of demnd	r_t	Nominal int rate	ε_{rt}	Inn. to r_t
4	ϕ	Index of price stckness	y_t	Nominal output		
5	Π	St. state level of infl.	π_t	Inflation		
6	ψ_π	Taylor rule infl. coeff.	c_t	Nominal cons.		
7	ψ_y	Taylor rule out. coeff.				
8	ρ_z	z_t persistence				
9	ρ_g	g_t persistence				
10	ρ_r	r_t persistence				
11	σ_z	ε_{zt} std error				
12	σ_g	ε_{gt} std error				
13	σ_r	ε_{rt} std error				
14	σ_{gz}	Covar of ε_{gt} and ε_{zt}				
15	σ_{rz}	Covar of ε_{rt} and ε_{zt}				
16	σ_{rg}	Covar of ε_{rt} and ε_{gt}				

Table 1: An and Schorfheide (2007) model parameter and variable names.

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - g_t) \tag{6}$$

$$c_t = y_t - g_t \tag{7}$$

where κ is a composite function of four underlying structural parameters.

$$\kappa = \tau \frac{1-\nu}{\nu\phi\Pi^2}$$

The 3×1 vector of innovations $\varepsilon_t = [\varepsilon_{zt}, \varepsilon_{gt}, \varepsilon_{rt}]'$ is iid mean-zero Gaussian with variance-covariance matrix

$$\Sigma_\varepsilon(\theta) \equiv E(\varepsilon_t \varepsilon_t'; \theta) = \begin{bmatrix} \sigma_z^2 & \cdot & \cdot \\ \sigma_{gz} & \sigma_g^2 & \cdot \\ \sigma_{rz} & \sigma_{rg} & \sigma_r^2 \end{bmatrix} \tag{8}$$

Notice, in this paper I allow for the possibility that the off-diagonal covariances are nonzero. θ is the 16×1 vector structural parameter

$$\underset{(16\times 1)}{\theta} = (\tau, \beta, \nu, \phi, \Pi, \psi_\pi, \psi_y, \rho_z, \rho_g, \rho_r, \sigma_z, \sigma_g, \sigma_r, \sigma_{gz}, \sigma_{rz}, \sigma_{rg})' \tag{9}$$

Respectively, Equations (2)-(7) are rule of motion for TFP an government spending, a Taylor rule, Euler equation, Phillips curve, and aggregate accounting equality. At first blush, the resemblance between this model and the VAR in Equation (1) is not obvious. In particular, very specific assumptions are necessary for a given DSGE model to have reduced form representation (See Ravenna (2007) and Giacomini (2013)). Furthermore, in order to form any reduced form representation, it is usually necessary to used a numerical solution algorithm like Sims (2002)'s. Such a black box makes it impossible to exploit the certain identifiability of the VAR parameters for the purposes of determining the identification of the structural parameters.

In the next section, I show how a simplified version of the An and Schorfheide model may be solved analytically, and that this solution is in fact a special case of Equation (1). Then, I do the same for the full model.

4.1 Simplified An and Schorfheide Model

Consider the special case in which TFP is iid, $z_t = \varepsilon_{zt}$, and government spending is zero, $g_t = 0 \; \forall \; t$. Then, the model has three equations,

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_\pi \pi_t + (1-\rho_r)\psi_y(y_t - g_t) + \varepsilon_{rt}$$

$$y_t = E_t y_{t+1} + g_t - E_t g_{t+1} - (1/\tau)(r_t - E_t \pi_{t+1}) + (1/\tau)\varepsilon_{zt}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - g_t)$$

The solution of this model implies the following rule of motion for interest rates:

$$r_t = \phi_{rr}(\theta) r_{t-1} + d_{rz}(\theta)\varepsilon_{zt} + d_{rg}(\theta)\varepsilon_{gt} + d_{rr}(\theta)\varepsilon_{rt}$$

Redefine $\rho_r = \phi_{rr}$ and $\varepsilon_{rt} = d_{rz}\varepsilon_{zt} + d_{rg}\varepsilon_{gt} + d_{rr}\varepsilon_{rt}$. In addition, let inflation be augmented by a shock, $\pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \varepsilon_{\pi t}$, and define $\varepsilon_{yt} = (1/\tau)\varepsilon_{zt}$. Thus, the above three equations become the following:[13]

[13]In other words, the reduced form rule of motion for interest rates, Equation (10) is equivalently the minimal state variable solution of a Taylor rule that responds to both inflation and output when r_{t-1} is

$$r_t = \rho_r r_{t-1} + \varepsilon_{rt} \tag{10}$$

$$y_t = E_t y_{t+1} - (1/\tau)\left(r_t - E_t \pi_{t+1}\right) + \varepsilon_{yt} \tag{11}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \varepsilon_{\pi t} \tag{12}$$

The three exogenous variables ε_{rt}, ε_{yt}, and $\varepsilon_{\pi t}$ are iid mean-zero Gaussian innovations. While ε_{rt} and ε_{yt} are respectively the idiosyncratic portion of monetary policy and a technological innovation, $\varepsilon_{\pi t}$ may be conceptualized as a cost-push shock or observational error, for instance (See Ireland (2004)). Allowing for the possibility that three innovations are mutually correlated, the 3×1 vector of innovations $\varepsilon_t = [\varepsilon_{rt}, \varepsilon_{yt}, \varepsilon_{\pi t}]'$ has variance-covariance matrix

$$\Sigma_\varepsilon(\theta) \equiv E(\varepsilon_t \varepsilon_t'; \theta) = \begin{bmatrix} \sigma_r^2 & \cdot & \cdot \\ \sigma_{yr} & \sigma_y^2 & \cdot \\ \sigma_{\pi r} & \sigma_{\pi y} & \sigma_\pi^2 \end{bmatrix} \tag{13}$$

where θ is the 10×1 vector structural parameter

$$\underset{(10\times 1)}{\theta} = (\tau, \beta, \kappa, \rho_r, \sigma_r, \sigma_y, \sigma_\pi, \sigma_{yr}, \sigma_{\pi r}, \sigma_{\pi y})'$$

This model must be solved to be analyzed empirically. Since r_t is the only lagged variable, the solution is of the form $E_t y_{t+1} = \phi_{yr}(\theta) r_t$ and $E_t \pi_{t+1} = \phi_{\pi r}(\theta) r_t$. By the method of undetermined coefficients, $\phi_{yr} = (\phi_{yr} - 1/\tau(1 - \phi_{\pi r}))\,\rho_r$ and $\phi_{\pi r} = (\rho_r \beta \phi_{\pi r} + \kappa \phi_{yr})$ (See Galí (2008)). So, the vector $Y_t = [r_t, y_t, \pi_t]'$ has restricted Gaussian VAR(1) representation $Y_t = \Phi(\theta) Y_{t-1} + U_t$. U_t is a 3×1 vector of mean-zero Gaussian innovations with covariance matrix $E(U_t U_t'; \theta) = \Omega(\theta)$.

the only lagged endogenous variable in the model; see also McCallum (1983) and McCallum (1999). Thus, Equation (10) is simply a reparameterization of the Taylor rule and this model will not produce sunspots.

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \rho_r & 0 & 0 \\ \phi_{yr} & 0 & 0 \\ \phi_{\pi r} & 0 & 0 \end{bmatrix}}_{\Phi(\theta)} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \tag{14}$$

$$\underbrace{\begin{bmatrix} \omega_r^2 & \cdot & \cdot \\ \omega_{yr} & \omega_y^2 & \cdot \\ \omega_{\pi r} & \omega_{\pi y} & \omega_\pi^2 \end{bmatrix}}_{\Omega(\theta)} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ \phi_{yr}/\rho_r & 1 & 0 \\ \phi_{\pi r}/\rho_r & \kappa & 1 \end{bmatrix}}_{D(\theta)} \underbrace{\begin{bmatrix} \sigma_r^2 & \cdot & \cdot \\ \sigma_{yr} & \sigma_y^2 & \cdot \\ \sigma_{\pi r} & \sigma_{\pi y} & \sigma_\pi^2 \end{bmatrix}}_{\Sigma_\varepsilon(\theta)} \underbrace{\begin{bmatrix} 1 & \phi_{yr}/\rho_r & \phi_{\pi r}/\rho_r \\ 0 & 1 & \kappa \\ 0 & 0 & 1 \end{bmatrix}}_{D(\theta)'} \tag{15}$$

where

$$\phi_{yr}(\theta) = \frac{1-\rho_r\beta}{\kappa} \times \phi_{\pi r}(\theta)$$

$$\phi_{\pi r}(\theta) = -\left(\frac{1}{\tau} \times \frac{\rho_r}{1-\rho_r}\right) \Big/ \left(\frac{1-\rho_r\beta}{\kappa} - \left(\frac{1}{\tau} \times \frac{\rho_r}{1-\rho_r}\right)\right)$$

The unique reduced form parameters are collected in the 9×1 vector reduced form parameter

$$\underset{(9\times 1)}{\Pi(\theta)} = (\rho_r, \phi_{yr}, \phi_{\pi r}, \omega_r, \omega_y, \omega_\pi, \omega_{yr}, \omega_{\pi r}, \omega_{\pi y})'$$

It will ultimately prove useful to discuss specific values of the structural parameters. Table 2 provides a candidate calibration θ_0, along with upper and lower bounds for the individual elements of θ; these make up the theoretically justifiable population parameter space Θ_0. One possible interpretation of Θ_0 is as the boundaries of diffuse priors over θ. Another is simply the support of any bounded prior, or the maximum admissible space. At θ_0,

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 0.75 & 0 & 0 \\ 1.55 & 0 & 0 \\ 2.04 & 0 & 0 \end{bmatrix}}_{\Phi(\theta_0)} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \qquad \Omega(\theta_0) = (1\text{e}-4) \times \begin{bmatrix} 4 & \cdot & \cdot \\ 9 & 25 & \cdot \\ 12 & 28 & 40 \end{bmatrix}$$

	Param	Lower	$\boldsymbol{\theta_0}$	Upper
1	τ	0.1	2	3.5
2	β	0.975	0.9975	1-ε
3	κ	ε	0.33	3
4	ρ_r	ε	0.75	1-ε
5	σ_r	ε	2e-2	1
6	σ_y	ε	2e-2	1
7	σ_π	ε	2e-2	1
8	σ_{yr}	-1	1e-4	1
9	$\sigma_{\pi r}$	-1	1e-4	1
10	$\sigma_{\pi y}$	-1	-1e-4	1

Table 2: Simplified An and Schorfheide model candidate calibration θ_0 and theoretically motivated parameter space Θ_0. ε =1e-6. Note, parameter bounds are purposefully allowed to be generous; see also Table 1 for microfounded definitions.

4.2 Full An and Schorfheide Model

While the simplified model may be solved analytically and has parsimonious reduced form representation, a natural question is whether the same is true of more empirically relevant specifications with latent state variables. It turns out that an exactly analogous solution methodology may be pursued using symbolic computation.

Recall that the structural parameter for the full model is the 16×1 vector given in Equation (9). While the solution of the simplified model followed from the fact that only lagged interest rates appeared in the equilibrium conditions, the same is not true here. To approach the solution of the full model analytically, note that the minimal solution of the model has the following general form (See Komunjer and Ng (2011) and Kailath et al. (2000)):

$$\underbrace{\begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{rz} & c_{rg} & c_{rr} \end{bmatrix}}_{A(\theta)} \underbrace{\begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d_{rz} & d_{rg} & d_{rr} \end{bmatrix}}_{B(\theta)} \underbrace{\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \end{bmatrix}}_{\varepsilon_t}$$

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} c_{rz} & c_{rg} & c_{rr} \\ c_{yz} & c_{yg} & c_{yr} \\ c_{\pi z} & c_{\pi g} & c_{\pi r} \end{bmatrix}}_{C(\theta)} \begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} d_{rz} & d_{rg} & d_{rr} \\ d_{yz} & d_{yg} & d_{yr} \\ d_{\pi z} & d_{\pi g} & d_{\pi r} \end{bmatrix}}_{D(\theta)} \begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \end{bmatrix}$$

where the scalars $c_{ij}(\theta)$ and $d_{ij}(\theta)$ are a-priori unknown functions of the structural parameters θ, while the sizes of the vectors of observables Y_t , states X_t, and innovations ε_t are $n_Y = n_X = n_\varepsilon = 3$. Given $E_t\varepsilon_{t+1} = 0_{3\times 1}$, the observation equation implies

$$E_t \underbrace{\begin{bmatrix} r_{t+1} \\ y_{t+1} \\ \pi_{t+1} \end{bmatrix}}_{Y_{t+1}} = \underbrace{\begin{bmatrix} c_{rz} & c_{rg} & c_{rr} \\ c_{yz} & c_{yg} & c_{yr} \\ c_{\pi z} & c_{\pi g} & c_{\pi r} \end{bmatrix}}_{C(\theta)} \begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix}$$

Plugging in the corresponding equations for $E_t y_{t+1}$ and $E_t \pi_{t+1}$, along with $E_t z_{t+1} = \rho_z z_t$, into aggregate demand, Equation (5),

$$y_t = \underbrace{\left(c_{yz} + \frac{1}{\tau}c_{\pi z} + \frac{\rho_z}{\tau}\right)}_{f_{yz}} z_t + \underbrace{\left(c_{yg} + \frac{1}{\tau}c_{\pi g} + (1-\rho_g)\right)}_{f_{yg}} g_t + \underbrace{\left(c_{yr} + \frac{1}{\tau}c_{\pi r} - \frac{1}{\tau}\right)}_{f_{yr}} r_t$$

Plugging the corresponding equation for $E_t \pi_{t+1}$, along with the expression for y_t just derived into the Phillips curve, Equation (6),

$$\pi_t = \underbrace{\left(\kappa c_{yz} + \left(\beta + \frac{\kappa}{\tau}\right) c_{\pi z} + \frac{\rho_z \kappa}{\tau}\right)}_{f_{\pi z}} z_t + \underbrace{\left(\kappa c_{yg} + \left(\beta + \frac{\kappa}{\tau}\right) - \rho_g \kappa\right)}_{f_{\pi g}} g_t + \underbrace{\left(\kappa c_{yr} + \left(\beta + \frac{\kappa}{\tau}\right) c_{\pi r} - \frac{\kappa}{\tau}\right)}_{f_{\pi r}} r_t$$

So, collecting the last two equations, and using the implicitly defined terms $f_{ij}(\theta)$,

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ f_{yz} & f_{yg} & f_{yr} \\ f_{\pi z} & f_{\pi g} & f_{\pi r} \end{bmatrix}}_{F(\theta)} \underbrace{\begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix}}_{X_t} \tag{16}$$

Finally, given ABCD representation and the fact that $Y_t = F(\theta)X_t$,

$$\underbrace{\begin{bmatrix} c_{rz} & c_{rg} & c_{rr} \\ c_{yz} & c_{yg} & c_{yr} \\ c_{\pi z} & c_{\pi g} & c_{\pi r} \end{bmatrix}}_{C(\theta)} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ f_{yz} & f_{yg} & f_{yr} \\ f_{\pi z} & f_{\pi g} & f_{\pi r} \end{bmatrix}}_{F(\theta)} \times \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{rz} & c_{rg} & c_{rr} \end{bmatrix}}_{A(\theta)} \tag{17}$$

Since the elements of F have been expressed in terms of the elements of C and θ, the system in Equation (17) yields 9 equations and 9 unknowns, the elements of C. Although it is infeasible to solve for these as functions of θ by hand, it is straightforward to make use of symbolic computation software for this purpose. Solving the model using `MATLAB`'s built-in `MuPAD` symbolic computation software reveals that there are exactly three solutions.[14] However, only one of these solutions implies a stable solution at θ_0, as judged by the modulus of the eigenvalues of $A(\theta_0)$. The generalized functional form of this unique stable solution is

$$C(\theta) = \begin{bmatrix} c_{rz}(\theta) & 0 & c_{rr}(\theta) \\ c_{yz}(\theta) & \rho_g & c_{yr}(\theta) \\ c_{\pi z}(\theta) & 0 & c_{\pi r}(\theta) \end{bmatrix}$$

where each of the 6 functions not explicitly shown are distinct functions of the structural parameters. Although their functional forms are extremely complicated and too unintuitive to be worth stating, they are closed-form. The two zeros are exactly zero.

Solving for $C(\theta)$ involved using the fact that $Y_t = F(\theta)X_t$ to infer that $C(\theta) = F(\theta)\times A(\theta)$. Notice that the same fact can be used to conclude that $D(\theta) = F(\theta)\times B(\theta)$. Taking the expressions for $C(\theta)$ as given and again utilizing symbolic computation,

$$\underbrace{\begin{bmatrix} d_{rz} & d_{rg} & d_{rr} \\ d_{yz} & d_{yg} & d_{yr} \\ d_{\pi z} & d_{\pi g} & d_{\pi r} \end{bmatrix}}_{D(\theta)} \times \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ 0 & 0 & \rho_r \end{bmatrix}}_{\rho(\theta)} = \underbrace{\begin{bmatrix} c_{rz} & 0 & c_{rr} \\ c_{yz} & \rho_g & c_{yr} \\ c_{\pi z} & 0 & c_{\pi r} \end{bmatrix}}_{C(\theta)}$$

[14] All `MATLAB and C code` is available on the author's website.

where ρ is a 3×3 matrix containing all, and only, the 3 persistence terms of the model, ρ_z, ρ_g, and ρ_r. So, using the simple formula for the inverse of a diagonal matrix, the matrices A, B, C, and D may be written exclusively in terms of $C(\theta)$ and $\rho(\theta)$.

$$\underbrace{\begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{rz} & 0 & c_{rr} \end{bmatrix}}_{A(\theta)} \underbrace{\begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c_{rz}/\rho_z & 0 & c_{rr}/\rho_r \end{bmatrix}}_{B(\theta)=A(\theta)\times\rho(\theta)^{-1}} \underbrace{\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \end{bmatrix}}_{\varepsilon_t}$$

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} c_{rz} & 0 & c_{rr} \\ c_{yz} & \rho_g & c_{yr} \\ c_{\pi z} & 0 & c_{\pi r} \end{bmatrix}}_{C(\theta)} \begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} c_{rz}/\rho_z & 0 & c_{rr}/\rho_r \\ c_{yz}/\rho_z & 1 & c_{yr}/\rho_r \\ c_{\pi z}/\rho_z & 0 & c_{\pi r}/\rho_r \end{bmatrix}}_{D(\theta)=C(\theta)\times\rho(\theta)^{-1}} \begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \end{bmatrix}$$

As in the simplified model, θ is assigned a candidate calibration θ_0, and parameter space Θ_0, in Table 3. It is verified that the values of this analytical ABCD solution correspond to Sims (2002)'s numerical `gensys.m` solution at the same point, up to only small approximation error inherent in the numerical solution (Refer to Komunjer and Ng (2011) Table I for these value). However, as previously discussed, the state space parameters are not reduced form parameters. Thus, it is desirable to explore any companion forms the model might have. It is now useful to state a simple result.

Reduced Form. *Let Y_t have ABCD representation. If there exists an invertible matrix $F(\theta)$ such that $Y_t = F(\theta)X_t$, then Y_t also has VAR(1) representation*

$$Y_t = CAC^{-1}Y_{t-1} + D\varepsilon_t$$

Proof. If Y_t has ABCD representation, but also $Y_t = FX_t$ for invertible F, then using the state equation, $FX_t = FAF^{-1}FX_t + FB\varepsilon_t$ and thus $Y_t = FAF^{-1}Y_{t-1} + FB\varepsilon_t$. Now using the observation equation, $C = FA$ and $D = FB$ exactly. Thus, $F = CA^{-1}$ so that $FAF^{-1} = CAC^{-1}$ and $FB = D$. □

Without any exogenous restrictions, the An and Schorfheide model satisfies $Y_t =$

	Param	Lower	$\boldsymbol{\theta_0}$	Upper
1	τ	0.1	2	3.5
2	β	0.975	0.9975	1-ε
3	ν	ε	0.1	1
4	ϕ	50	53.68	60
5	Π	1+ε	1.008	1.03
6	ψ_π	-1	1.5	3
7	ψ_y	-1	0.125	1.25
8	ρ_z	ε	0.9	1-ε
9	ρ_g	ε	0.95	1-ε
10	ρ_r	ε	0.75	1-ε
11	σ_z	ε	3e-2	1
12	σ_g	ε	6e-2	1
13	σ_r	ε	2e-2	1
14	σ_{gz}	-1	1e-4	1
15	σ_{rz}	-1	1e-4	1
16	σ_{rg}	-1	-1e-4	1

Table 3: Full An and Schorfheide (2007) candidate calibration θ_0 and space Θ_0. ε=1e-6.

$F(\theta)X_t$ from Equation (16). Thus, its observables Y_t have restricted VAR(1) representation. Defining U_t to be a 3×1 vector of nonstructural innovations $D\varepsilon_t$ with covariance matrix $E(U_tU_t') \equiv \Omega(\theta)$, by the above result, the observables Y_t have the following VAR(1) reduced form representation, another special case of Equation (1):

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \phi_{rr} & 0 & \phi_{r\pi} \\ \phi_{yr} & \rho_g & \phi_{y\pi} \\ \phi_{\pi r} & 0 & \phi_{\pi\pi} \end{bmatrix}}_{\Phi(\theta)} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \tag{18}$$

Given that $D(\theta) = C(\theta) \times \rho(\theta)^{-1}$ and that $\rho(\theta)^{-1'} = \rho(\theta)^{-1}$ since $\rho(\theta)$ is diagonal,

$$\underbrace{\begin{bmatrix} \phi_{rr} & 0 & \phi_{r\pi} \\ \phi_{yr} & \rho_g & \phi_{y\pi} \\ \phi_{\pi r} & 0 & \phi_{\pi\pi} \end{bmatrix}}_{\Phi(\theta)} = \underbrace{\begin{bmatrix} c_{rz} & 0 & c_{rr} \\ c_{yz} & \rho_g & c_{yr} \\ c_{\pi z} & 0 & c_{\pi r} \end{bmatrix}}_{C(\theta)} \times \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{rz} & 0 & c_{rr} \end{bmatrix}}_{A(\theta)} \times \underbrace{\begin{bmatrix} c_{rz} & 0 & c_{rr} \\ c_{yz} & \rho_g & c_{yr} \\ c_{\pi z} & 0 & c_{\pi r} \end{bmatrix}^{-1}}_{C(\theta)^{-1}} \tag{19}$$

$$\underbrace{\begin{bmatrix} \omega_r^2 & \cdot & \cdot \\ \omega_{yr} & \omega_y^2 & \cdot \\ \omega_{\pi r} & \omega_{\pi y} & \omega_\pi^2 \end{bmatrix}}_{\Omega(\theta)} = \underbrace{\begin{bmatrix} c_{rz} & 0 & c_{rr} \\ c_{yz} & \rho_g & c_{yr} \\ c_{\pi z} & 0 & c_{\pi r} \end{bmatrix}}_{C(\theta)} \times \underbrace{\left[\begin{array}{c|c|c} \frac{\sigma_z^2}{\rho_z^2} & \cdot & \cdot \\ \hline \frac{\sigma_{gz}}{\rho_z \rho_g} & \frac{\sigma_g^2}{\rho_g^2} & \cdot \\ \hline \frac{\sigma_{rz}}{\rho_r \rho_z} & \frac{\sigma_{rg}}{\rho_r \rho_g} & \frac{\sigma_r^2}{\rho_r^2} \end{array}\right]}_{\rho(\theta)^{-1} \times \Sigma_\varepsilon(\theta) \times \rho(\theta)^{-1}} \times \underbrace{\begin{bmatrix} c_{rz} & c_{yz} & c_{\pi z} \\ 0 & \rho_g & 0 \\ c_{rr} & c_{yr} & c_{\pi r} \end{bmatrix}}_{C(\theta)'} \tag{20}$$

The unique reduced form parameters are collected in the 13×1 vector parameter

$$\underset{(13\times 1)}{\Pi(\theta)} = (\phi_{rr}, \phi_{yr}, \phi_{\pi r}, \rho_g, \phi_{r\pi}, \phi_{y\pi}, \phi_{\pi\pi}, \omega_r, \omega_y, \omega_\pi, \omega_{yr}, \omega_{\pi r}, \omega_{\pi y})'$$

At θ_0, Equation (18) has the following realization:

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 0.79 & 0 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0 & 0.62 \end{bmatrix}}_{\Phi(\theta_0)} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \qquad \Omega(\theta_0) = (1\mathrm{e}-4) \times \begin{bmatrix} 6 & \cdot & \cdot \\ 7 & 58 & \cdot \\ 7 & 21 & 20 \end{bmatrix}$$

Reduced form representation of the model with-means is also easy to derive, and available in the online `supplementary material`.

5 Identification Analysis

As motivated previously, the global identification of the reduced form VAR(1) Equation (1) is trivial, and furthermore, both the simplified and full versions of the model are now known to have restricted VAR(1) representation. Yet, the identification of θ is the relevant issue. In this section I study the identifiability of the structural parameters θ from the reduced form parameters Π.

5.1 Simplified Model

In order to distinguish whether θ is globally identified in the real plane, it is sufficient to show that θ is uniquely recoverable from Π regardless of θ's realization in the reals. Although the mapping from structural parameters θ to reduced form parameters Π is

	θ	1	2	3	4	5	6	7	8	9	10
Π		τ	β	κ	ρ_r	σ_r	σ_y	σ_π	σ_{yr}	$\sigma_{\pi r}$	$\sigma_{\pi y}$
1	ρ_r				×						
2	ϕ_{yr}	×	×	×	×						
3	$\phi_{\pi r}$	×	×	×	×						
4	ω_r					×					
5	ω_y	×	×	×	×	×	×		×		
6	ω_π	×	×	×	×	×	×	×	×	×	×
7	ω_{yr}	×	×	×	×	×			×		
8	$\omega_{\pi r}$	×	×	×	×	×			×	×	
9	$\omega_{\pi y}$	×	×	×	×	×	×		×	×	×

Table 4: Simplified model functional mapping $g : \theta \to \Pi$.

nonlinear (See Equations (14) and (15)), by virtue of the analytical solution developed in the preceding section, the correspondence is closed-form. In place of the complicated functions themselves, ×'s are used in Table 4 to represent the dependence of each element of Π on each element of θ. For instance, the standard error ω_π is a function of all 10 structural parameters. As noted in the figure, it will henceforth clarify the analysis to name this functional dependence $\Pi = g(\theta)$. A similar identification table was also recently utilized by Hamilton and Wu (2012) to investigate the identification of affine term structure models.

The number of structural parameters, $n_\theta = 10$ columns, exceeds the number of reduced form parameters, $n_\Pi = 9$ rows; in other words, it is immediately apparent that the necessary order condition for identification, $n_\Pi \geq n_\theta$, is violated. Thus, at least one of the structural parameters must be set to a constant for the complement subset to be conditionally identified. In order to distinguish which structural parameter is such a candidate, or if more than one parameter must be set, one possible procedure is to check local identification. As motivated by Iskrev (2010), a reasonable starting point is to compute the Jacobian $J(\theta_0) = \partial\Pi/\partial\theta'|_{\theta=\theta_0}$ and successively eliminate columns until those that remain are full column rank; the dropped columns correspond to parameters to be set. However, such an approach is only valid at θ_0 and overlooks the issue of global identification more broadly.

Instead, consider the following logical points: First, if setting one structural parameter did in fact result in the remaining 9 being conditionally locally identified, those 9

would be exactly locally identified by the 9 elements of Π. Let ϑ be such a hypothetical 9×1 subvector of θ, and $\overline{\alpha}$ the scalar fixed parameter. Second, a value of ϑ which yields Π_0 conditional on $\overline{\alpha}$ is exactly ϑ_0 such that $\Pi_0 = g(\vartheta_0; \overline{\alpha})$. Using Corollary 1, ϑ is globally identified at ϑ_0 if and only if there exists no $\vartheta_0^* \neq \vartheta_0$ such that $\Pi_0 = g(\vartheta_0^*; \overline{\alpha})$. Third, typically, one would have to search for such a ϑ_0^* numerically, a method which is subject to the dilemma of false positives. However, since $n_\vartheta = 9 = n_\Pi$ and the solution is analytical, a simple non-numerical method is to simply check whether a unique inverse of the vector-valued function g exists. When it does, $\vartheta_0 = g^{-1}(\Pi_0; \overline{\alpha})$.

So, ϑ is globally identifiable at any given point in any parameter space if $n_\vartheta = n_\Pi$ and g is injective. The latter is verified if g^{-1} exists and is unique. Thus, the question is whether a unique inverse of the specific functional form of g for the O model exists. First, by examining the functional form of the vector-valued function $g(\theta)$ represented by Table 4, it appears that the parameter σ_π is a reasonable candidate for the role of $\overline{\alpha}$. This implies that the parameter to be analyzed using the analytical approach is

$$\underset{(9\times 1)}{\vartheta} = (\tau, \beta, \kappa, \rho_r, \sigma_r, \sigma_y, \sigma_\pi, \sigma_{yr}, \sigma_{\pi r}, \sigma_{\pi y})'$$

Using the analytical solution to invert the mapping implies that each of the 5 parameters τ, ρ_r, σ_r, σ_y, and σ_{yr} have unique functional forms in terms of the vector reduced form parameter. For example, given a realization Π, the unique expression for τ is

$$\tau = -\frac{\rho_r}{1 - \rho_r} \times \frac{1 - \phi_{\pi r}}{\phi_{yr}}$$

and ρ_r is itself one of the reduced form parameters. The remaining expressions for σ_r, $\sigma_{\pi r}$, and σ_{yr} are closed-form, but too complicated to provide any intuition. The most important result, however, has to do with the remaining 4 elements of ϑ. Even after normalizing the relationship between standard errors and variances, for any Π, there are exactly 2 expressions for each of β, κ, $\sigma_{\pi r}$, and $\sigma_{\pi y}$ in terms of Π, no more, no less.

		$\boldsymbol{g(x;\overline{\sigma}_\pi)}$		$\boldsymbol{\vartheta_0}$	$\boldsymbol{\vartheta_0^*}$
1	ρ_r	0.75	τ	2	2
2	ϕ_{yr}	1.55	$\boldsymbol{\beta}$	0.9975	**1.51**
3	$\phi_{\pi r}$	2.04	$\boldsymbol{\kappa}$	0.33	**-0.17**
4	ω_r	2e-2	ρ_r	0.75	0.75
5	ω_y	5e-2	σ_r	2e-2	2e-2
6	ω_π	6e-2	σ_y	2e-2	2e-2
7	ω_{yr}	9e-4	σ_{yr}	1e-4	1e-4
8	$\omega_{\pi r}$	12e-4	$\boldsymbol{\sigma_{\pi r}}$	1e-4	**2e-4**
9	$\omega_{\pi y}$	28e-4	$\boldsymbol{\sigma_{\pi y}}$	-1e-4	**1e-4**

Table 5: Simplified model observational equivalence in $\mathbb{R}^9$ but not Θ_0. $\Pi = g(x;\overline{\sigma}_\pi)$; $x = \vartheta_0$ or ϑ_0^*.

In other words, g is *not* injective, and its inverse yields two values:

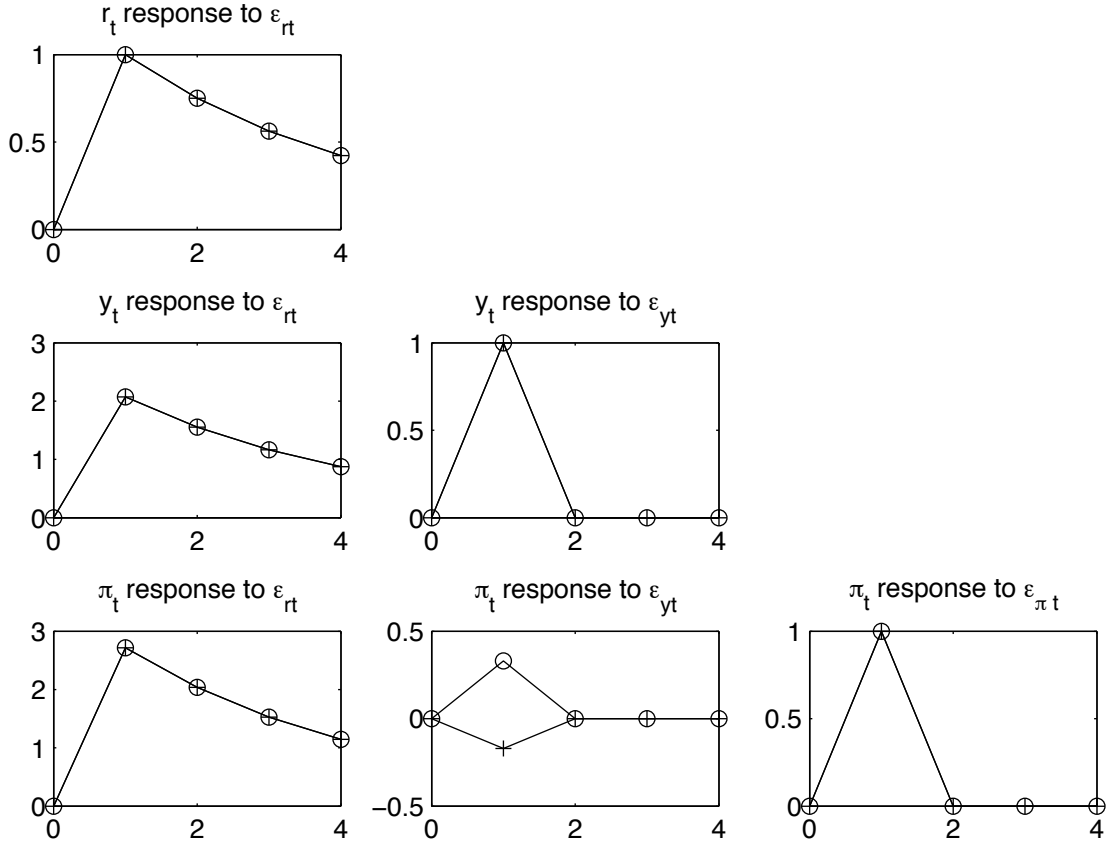
$$g^{-1}(g(\vartheta_0;\overline{\sigma}_\pi);\overline{\sigma}_\pi) \begin{cases} \nearrow \vartheta_0 \\ \searrow \vartheta_0^* \end{cases}$$

The term *inverse* is technically a misnomer since g^{-1} is multiple-valued. For example, at ϑ_0 the value corresponding to θ_0 listed in Table 2, there exists another value, $\vartheta_0^* \in \mathbb{R}^9$ but $\vartheta_0^* \notin \Theta_0$, yielding the same Π at $\overline{\sigma}_\pi = 2\text{e}-2$. See Table 5.

There are several important consequences of Table 5. First, ϑ_0 and ϑ_0^* are observationally equivalent despite that the 9×9 Jacobian $J(\vartheta_0;\overline{\sigma}_\pi) = \partial\Pi(\vartheta;\overline{\sigma}_\pi)/\partial\vartheta'|_{\vartheta=\vartheta_0}$ is full rank, a necessary and sufficient condition for local identification. Second, the value ϑ_0^* – and specifically, the values of β and κ at ϑ_0^* – fall outside of Θ_0. Therefore, an assessment of global identification at ϑ_0 based on numerically searching over Θ_0 would conclude simply that ϑ is globally identified in Θ_0, whereas this conclusion is possibly misleading of the fact that there is an observationally equivalent ϑ_0^* in $\mathbb{R}^9$ outside of, but relatively close to, Θ_0. This is important since, as will be shown, small sample biases may cause values of estimators to fall outside of theoretically feasible regions; I will return to the theoretical feasibility of $\beta > 1$ and $\kappa < 0$ momentarily. Third, computational time is a fraction of a second.

A natural next question is whether the economic implications of both ϑ_0 and ϑ_0^* are

Figure 2: Simplified model impulse-responses for observationally equivalent points: ϑ_0 ($\circ$) vs ϑ_0^* ($+$). Triangularity of impulse-responses reflects triangularity of $D(\theta)$ matrix in Equation (15).



the same. I provide impulse-responses for both points in Figure 2. When the crosshairs corresponding to ϑ_0^* exactly hit the bullseyes corresponding to ϑ_0, the economic implications are equivalent. In only one case, π_t's response to ε_{yt}, do the economic implications of each point differ. In fact, this can be traced back to Equation (15) as arising from the difference in κ between points. There are two important implications of this observation. First, the monetary policy impulse-responses in the first column of Figure 2 are robust to observational equivalence. This means that if those impulse responses were the only object of economic interest, a valid normalization is to simply drop ϑ_0^*. Second, the economic implications of the two points for π_t's response to ε_{yt} differ. Therefore, if this is the object of economic analysis, simply dropping ϑ_0^* is not a normalization as has been defined in Definition 4, but something stronger.

The good news is that is reasonable to distinguish between ϑ_0 and ϑ_0^* on the basis of the bounded priors embodied in Θ_0. For instance, not only is $\kappa < 0$ contrary to typical theory, but $\beta > 1$ is certainly disconcerting. Specifically, if the space in which global identification is important is Θ_0, then Table 5 has simply shown us that ϑ_0 is globally identified in Θ_0. Thus, although this model does engender observational equivalencies, the two facts that global identification is defined in terms of a parameter space, and that macroeconomists have a good idea of what parameter spaces are important, allow the analyst to eliminate the nuisance ϑ_0^* in this case. It remains important, however, to realize that this is not necessarily a normalization with respect to all impulse responses as defined in Definition 4, but something stronger.

Yet, it is not always possible to distinguish between observationally equivalent points on the basis of the bounds of Θ_0 alone. Since evaluating global identification for a point may be done very efficiently, I am able to search the parameter space Θ_0 for values ϑ_1 which have an observationally equivalent ϑ_1^* that is also in Θ_0. Such an example is given in Table 6. The impulse responses for each point are given in Figure 3. Again, only π_t's response to ε_{yt} differs from point to point. However, in this case, both impulse responses have the same sign, owing to the fact that the value for κ at each point is the same. Evidently, observational equivalence is a more trying issue at the realization ϑ_1, since neither it nor ϑ_1^* may be easily eliminated. So, more elaborate

		$\boldsymbol{g(x;\overline{\sigma}_\pi)}$		$\boldsymbol{\vartheta_1}$	$\boldsymbol{\vartheta_1^*}$
1	ρ_r	0.93	τ	1.01	1.01
2	ϕ_{yr}	0.03	$\boldsymbol{\beta}$	0.9836	**0.9980**
3	$\phi_{\pi r}$	1.002	$\boldsymbol{\kappa}$	2.70	**2.25**
4	ω_r	0.41	ρ_r	0.93	0.93
5	ω_y	0.08	σ_r	0.41	0.41
6	ω_π	0.60	σ_y	0.07	0.07
7	ω_{yr}	0.03	σ_{yr}	0.03	0.03
8	$\omega_{\pi r}$	0.25	$\boldsymbol{\sigma_{\pi r}}$	-6.1e-3	**5.1e3**
9	$\omega_{\pi y}$	0.05	$\boldsymbol{\sigma_{\pi y}}$	-9e-4	**9e-4**

Table 6: Simplified model observational equivalence in Θ_0. $\Pi = g(x;\overline{\sigma}_\pi)$; $x = \vartheta_0$ or ϑ_0^*.

	Yes	No
Locally Identified	100 %	0 %
Globally Identified in Θ_0	65.76 %	34.24 %
Globally Identified in All $\Theta \subset \mathbb{R}^9$	0.04%	99.96 %

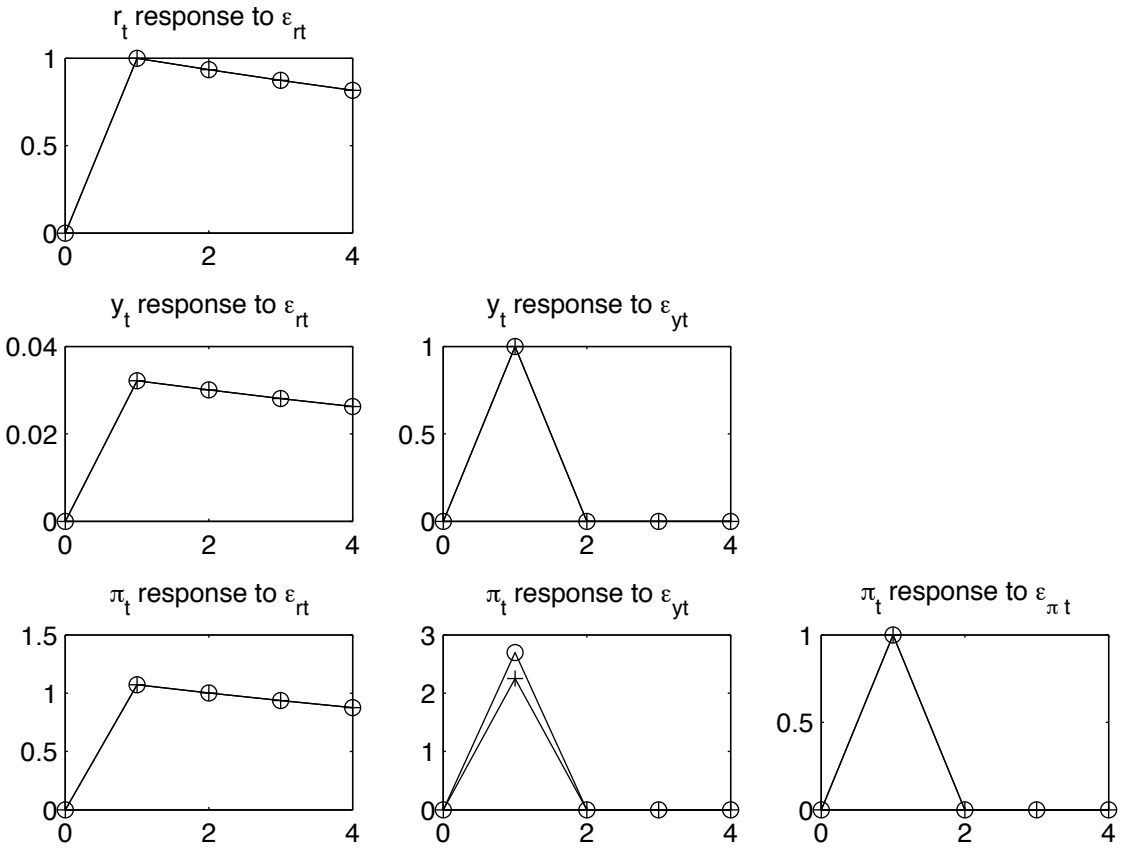
Table 7: Identification of ϑ in O model: 100,000 uniformly drawn points from Θ_0.

identifying restrictions are necessary.

Consider, then, the analyst who decides to simply choose ϑ_1^*, since the value of β is closer to the original value in θ_0 of 0.9975. As is well-known, many macroeconomists have very strong convictions about the value of this parameter in particular. This action would reflect, for example, the will of an analyst with a very tight prior for β centered at 0.9975. In fact, this ad-hoc identifying procedure is just as valid a normalization as any according to Definition 4, as long as the economic implications of interest are the same for both points. Thus, depending on the analyst's ultimate inferential goal, many possible identifying restrictions are valid normalizations. However, the analyst must also be careful to account for the fact that the identifying restrictions she uses have a direct effect on confidence intervals. This will be revisited shortly.

What is the likeliness of drawing a point like ϑ_0 versus a point like ϑ_1 in the parameter space Θ_0? Obviously, if there are relatively more ϑ_0, this is all else equal a good thing, since the analyst will be required to enforce relatively less stringent identifying restrictions. Table 7 presents the results of testing for local identification, and global identification in both Θ_0 and all $\Theta \subset \mathbb{R}^9$, at 100,000 uniformly drawn points from Θ_0. The only restriction I make on the points drawn is that the variance-

Figure 3: Simplified model impulse-responses for two theoretically plausible and observationally equivalent points: ϑ_1 ($\circ$) vs ϑ_1^* ($+$).



covariance matrix $\Sigma_\varepsilon(\vartheta;\overline{\alpha})$ must be positive semidefinite. Note, since $\Omega = D\Sigma_\varepsilon D'$ in Equation (15), Σ_ε positive semidefinite implies that so is Ω (See Abadir and Magnus (2005)). ϑ is locally identified at all points, but has a real-valued observationally equivalent point at all but 0.04% of draws.[15] Furthermore, the identifying restriction that ϑ must be in Θ_0 does not successfully yield a unique point in roughly a third of the space. Thus, stronger normalizations like the β criterion are frequently needed.

5.2 Plug-In Maximum Likelihood

While we are now aware of what types of normalizations are necessary, and when they must be implemented, we are not fully cognizant of the effects of these normalizations on confidence intervals. One approach to understand the statistical outcomes of a given normalization is to implement these identifying restrictions in the Monte Carlo distribution of a given estimator, such as the maximum likelihood estimator.[16] However, under normal circumstances, it would be impossible to construct such a distribution for DSGE models in particular. The reasons are three-fold. Firstly, without the knowledge of how many maxima are in $\mathbb{R}^9$, it would be impossible to determine how many maximum likelihood estimators exist for each draw i. Secondly, as demonstrated by Andreasen (2010), even the most sophisticated global algorithms for likelihood maximization, including simulated annealing and genetic numerical search, are prone to failure. Furthermore, search algorithms are only as accurate as the termination tolerance chosen by the analyst, which is particularly worrying given the topological characterization of weak identification is flatness in the likelihood surface (See Canova and Sala (2009)). Finally, likelihood maximization is computationally costly, particularly for the most reliable genetic algorithm, making it infeasible to compute the distribution for large N.

In fact, the mapping g^{-1} has other uses besides assessing identification; it may also be used as the basis of a "plug-in" maximum likelihood estimator. This tool

[15] All of these 0.04% of points have observationally equivalent points that are real, but yield Σ_ε which are not positive semidefinite, and thus not variance-covariance matrices.

[16] As in Stock et al. (2002), the Monte Carlo is also a natural place to begin analysis of weak identification.

is uniquely suited for computing the Monte Carlo distribution, and investigating the statistical consequences of implemented normalizations. I motivate this estimator using the minimum chi-squared estimator, or MCSE, implemented by Hamilton and Wu (2012) and first suggested by Rothenberg (1973). When $\widehat{\Pi}$ is an efficient estimator of the reduced form parameters, and $\widehat{\mathcal{I}}$ is an efficient estimator of the information matrix with respect to Π, the MCSE is defined by the following for conditional $\overline{\alpha}$:

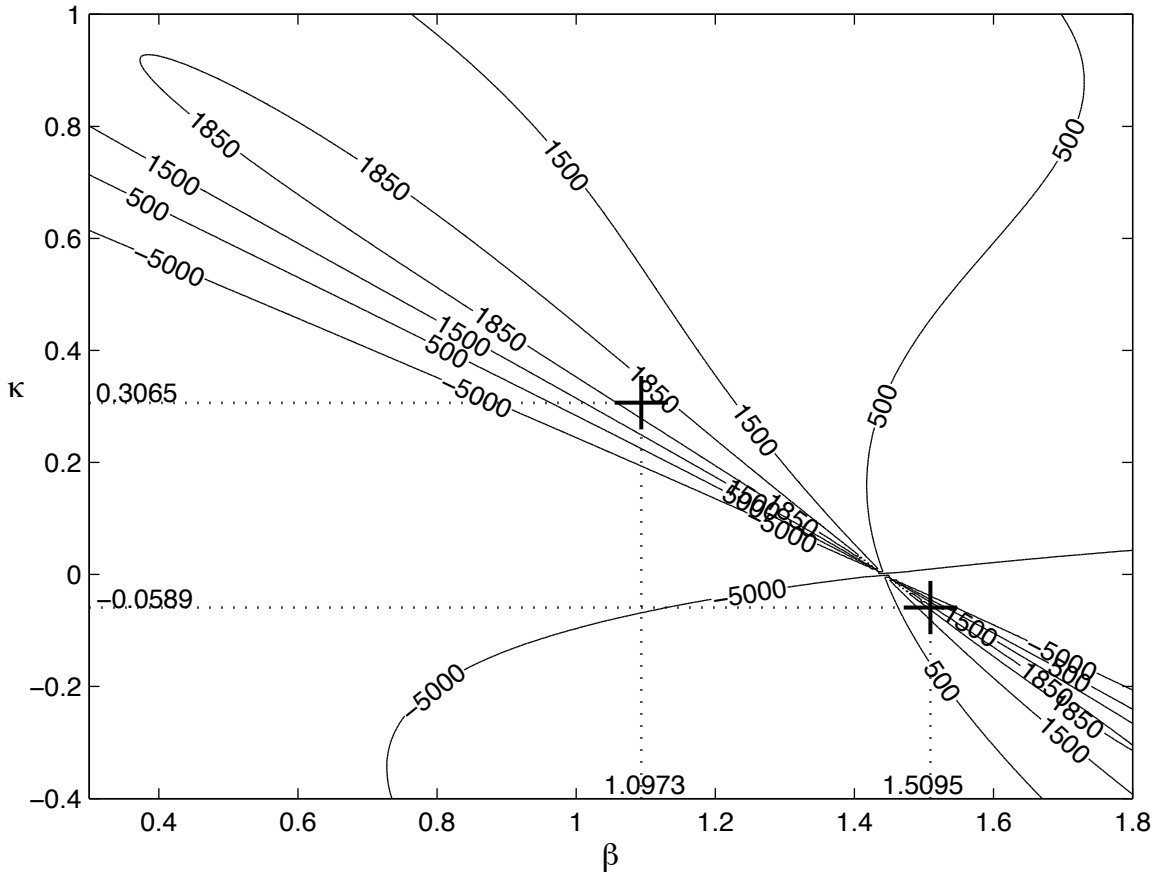
$$\widehat{\vartheta}_{MCSE} = \underset{\vartheta \in \Theta}{\arg\min}\ (\widehat{\Pi} - g(\vartheta; \overline{\alpha}))' \widehat{\mathcal{I}} (\widehat{\Pi} - g(\vartheta; \overline{\alpha}))$$

Note, since the models considered here are Gaussian, an efficient estimator of Π is restricted ordinary least squares, with the number of exclusion restrictions implied by the zeros in Φ. As Hamilton and Wu (2012) go on to show, fist, this estimator is asymptotically efficient. Second, in the special case that $n_\vartheta = n_\Pi$, the minimal value of the criterion is zero, in which case the analyst may simply use an identity matrix as the weighting matrix, and find the MCSE as equivalently the minimizer of $(\widehat{\Pi} - g(\vartheta; \overline{\alpha}))'(\widehat{\Pi} - g(\vartheta; \overline{\alpha}))$. Furthermore, in this special case, $\widehat{\vartheta}_{MCSE} = \widehat{\vartheta}_{MLE}$ exactly. But note, since this a quadratic form, $\widehat{\vartheta}_{MLE}$ is equivalently defined by $\widehat{\Pi} = g(\widehat{\vartheta}_{MLE}; \overline{\alpha})$. In other words, given any data set, under the conditional identification scheme $\overline{\alpha} = \overline{\sigma}_\pi$ there are exactly two points which maximize the likelihood:

$$g^{-1}(\widehat{\Pi}; \overline{\sigma}_\pi) \begin{cases} \nearrow \widehat{\vartheta} \\ \searrow \widehat{\vartheta}^* \end{cases}$$

Thus, by simply estimating $\widehat{\Pi}$ by restricted OLS, and given g^{-1} is known in closed-form, the analyst need only plug-in to obtain both points $\widehat{\vartheta}$ and $\widehat{\vartheta}^*$ that maximize the likelihood. As in Hamilton and Wu (2012), there is no uncertainty that each of $\widehat{\vartheta}$ and $\widehat{\vartheta}^*$ maximizes the likelihood. My unique extension of deriving g^{-1} directly also implies there is no uncertainty about the number of likelihood maxima or the reliability of search algorithms, nor is there computational toll from numerical search. To verify, I use the value ϑ_0 to generate a data set of length $T = 250$ and reestimate the

Figure 4: Contour plot of log-likelihood. Maximum achieved at identical 1909.9 for exactly two points +, which are the plug-in MLEs. Compare with depicted bimodal univariate likelihood in Figure 1.



parameters of the model using the plug-in technique. As predicted previously in Table (5), for instance, this estimator yields two different values for β and κ. By computing the contour plot of the likelihood over these values, it shown that the plug-in MLEs do maximize the likelihood function, with exactly equal values at both points.

5.3 Small Sample Distribution of MLE

It is now feasible to compute a Monte Carlo simulation of the small-sample standard errors of the MLE.[17] First, I consider the model subject to no identifying restrictions

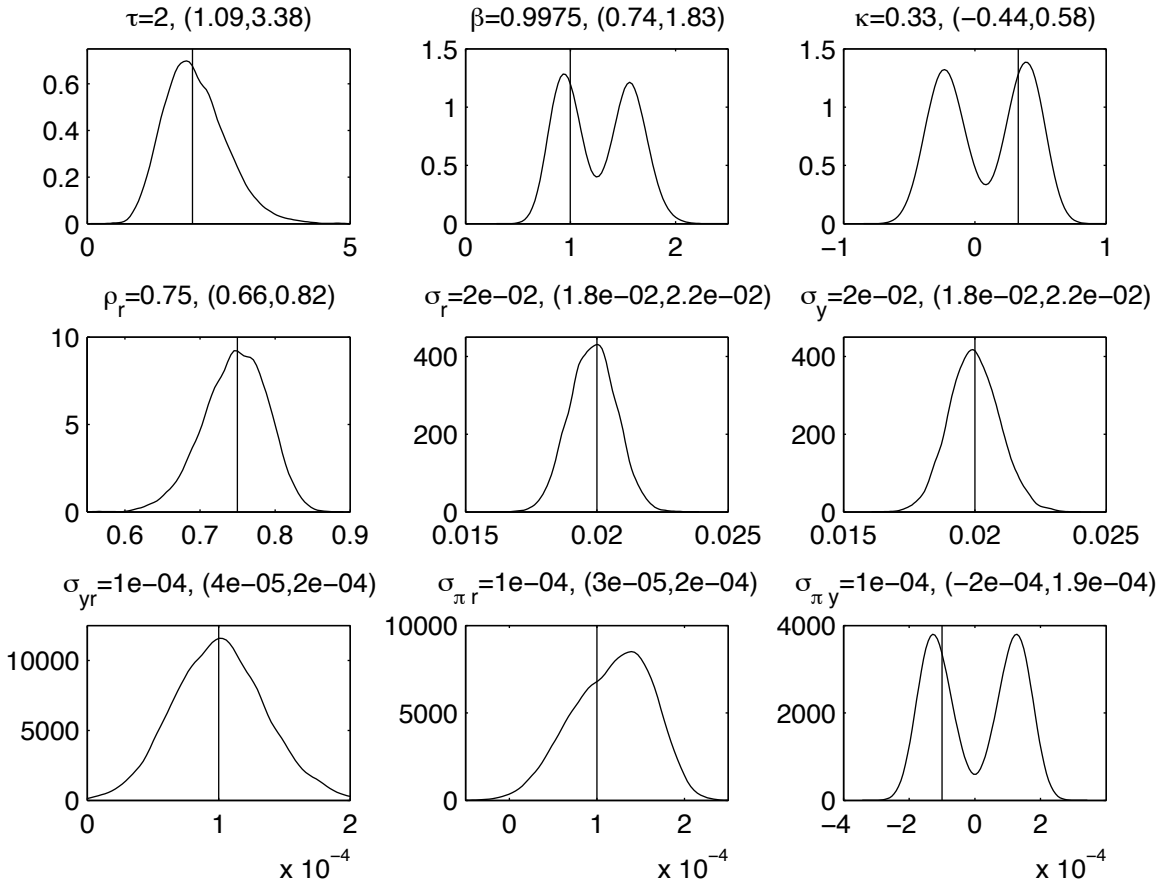
[17]Qu (2013), Guerron-Quintana et al. (2013), Dufour et al. (2013), and Andrews and Mikusheva (2013) have all discussed the issue of weak identification, and how to compute correctly sized large-sample confidence intervals and tests. Since the distributions I consider are bootstrapped small sample standard errors, they do not rely on the asymptotic approximations that break down under weak identification (See Stock et al. (2002), for example). Note, it is not generally possible to compute a bootstrapped distribution for weakly

is given in Figure 5 with $T = 250$ and N=10,00. This T is equivalent to 62.5 years of quarterly data, approximately the available window of post-war data on interest rates, output, and inflation. The entire distribution is computed in a few seconds. To understand how this distribution is computed requires first an understanding of the sample. While roughly 78% of the draws $\widehat{\Pi}_{\mathtt{i}}$ yield two real-valued estimators $\widehat{\vartheta}_{\mathtt{i}}$ and $\widehat{\vartheta}^{*}_{\mathtt{i}}$, 22% yield two complex-valued estimators, and 0% yield only one real-valued point. There are two conflicting perspectives on how to deal with the complex-valued estimators. One is to include these in the Monte Carlo distribution. The other is to simply try to find a local maximum in the reals. However, whereas the first approach takes this analysis out of the comfort zone of meaningful macroeconomic analysis, the second suggests an estimator known not to maximize the likelihood. Since there is no formal econometric basis for distinguishing between these two options, for the 22% of draws yielding two complex points, I take the safest route possible and simply say the estimator does not exist. Thus, the restricted estimator is technically a restricted MLE. When there are two real-valued estimators, each receives a weight of one in the distribution.

The multi-modal character of the likelihood in the parameters β, κ, and $\sigma_{\pi y}$, also tellingly reminiscent of the distribution of weak instrumental variables computed by Nelson and Startz (1990), is obvious. Moreover, there are a few other more subtle observations regarding the MLE distribution over the remaining parameters. Firstly, the parameter $\sigma_{\pi r}$ does not seem to have a bimodal distribution, even though it was the fourth structural parameter, besides β, κ, and $\sigma_{\pi y}$, to exhibit observational equivalence in Table 5. The reason is that the two modes pile-up immediately next to one another so there appears to be only one rightward-leading peak. Secondly, even single-modal parameters like the coefficient of relative risk aversion $\tau = -\rho_r/(1-\rho_r)\times(1-\phi_{\pi r})/\phi_{yr}$ seems to be biased (leftward). Evidently, even using the longest possible sample of post-war data, there is small sample bias in the MLE.

identified models. This is due to the fact that the bootstrap is known to be invalid under weak identification, an outcome related to the failure of Edgeworth expansion in this case (See Hall (1992)). However, the bootstrap statistics I calculate are applied only to the restricted OLS estimator of the reduced form VAR parameters, which are strongly identified.

Figure 5: Monte Carlo distribution of O Model restricted MLE at ϑ_0 in the real-valued codomain: T=250, N=10,000. Two-sided $(\alpha/2, 1-\alpha/2)$ confidence interval, $\alpha = 5\%$.



Pseudo code:

---

1. For draw $\widehat{\Pi}_{\mathtt{i}}$, compute both estimators $\widehat{\vartheta}_{\mathtt{i}}$ and $\widehat{\vartheta}^{*}_{\mathtt{i}}$ by plugging in to g^{-1}.
2. If both are outside of reals, the estimator is said to not exist for the given draw (22% of utilized sample).
3. If both estimators are in the reals, plot both in the Monte Carlo with equal weight 1.
4. If one is in the reals and one is outside, plot only the one inside with weight of 2* (Note, this option has no effect in the given sample).

---

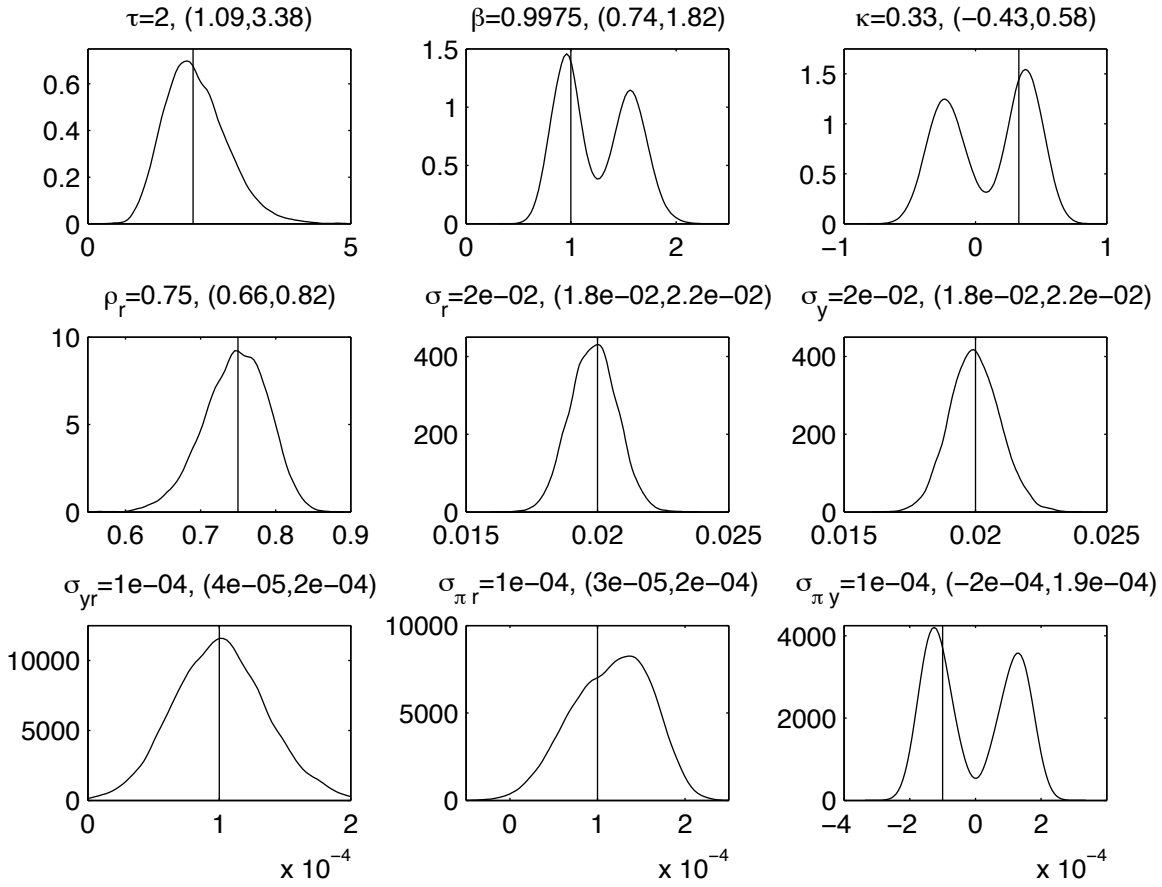
*A weight of 2 indicates that when one of the two estimators for draw `i` can be eliminated, the only remaining estimator receives double relative weight. This ensures that the small sample distribution is correctly computed across the sample of N draws. An alternative way to conceptualize this is that when there are two permissible estimators for a given draw `i`, the analyst has two half-weighted conflicting estimates. One reasonable story is that as the result of numerical issues, only one estimator is found at each draw, with equal probability.

Yet, the distribution of the maximum likelihood estimator itself is not the only important distribution to consider. As first explicitly shown by Hamilton et al. (2007), normalizations meant to achieve global identification can have unintended consequences on the properties of confidence intervals. Therefore, if the analyst is using a given normalization scheme to identify a given parameter value, it is critical that this identification scheme be embodied in the construction of the Monte Carlo distribution. First, consider again the analyst that uses bounded priors embodied by Θ_0 as the means for identification; specifically, a parameter is identified in Θ_0 if its observationally equivalent point is outside of this space. Recall, this normalization was a successful identifying restriction for ϑ_0 in Table 5. If two estimators are both inside of or both outside of Θ_0, there is no reasonable basis to distinguish between the two. Rather, there is only a unique estimator when just one is in Θ_0. The distribution of this estimator is given in Figure 6. In this case, the distributions are close to the original MLE, but now the second peaks for β, κ, and $\sigma_{\pi y}$ have shrunk slightly, reflecting the analyst's ability to distinguish between two points when one is outside of Θ_0. Note the standard errors for both β and κ have now changed, indicating a modest but realized affect of the identifying restrictions on the sizes of the two modes.

Now, consider an analyst confronted with a point like ϑ_1 in Table 6 for which the normalization that $\vartheta \in \Theta_0$ is not a successful identifying restriction. As was suggested, a normalization that works is to choose the value that yields a β closer to 0.9975, but certainly not larger than 1. The distribution in Figure 7 now implies confidence intervals which are notably different from the first MLE without identifying restrictions. In particular, the mode corresponding to the nuisance estimator $\widehat{\vartheta}^*$ has been leveled. Thus, any number of identifying restrictions useful, but confidence intervals are directly affected; the plug-in MLE allows the analyst to directly account for this discrepancy.

In this section, I have demonstrated the entirety of my approach to global identification, and accounting for the statistical consequences of necessary normalizations, using the O model. Next, I consider the L model, which may be studied similarly.

Figure 6: Monte Carlo distribution of Simplified Model MLE at ϑ_0 when the bounds of Θ_0 are used as the identifying restrictions: T=250, N=10,000. Two-sided $(\alpha/2, 1-\alpha/2)$ confidence interval, $\alpha = 5\%$.

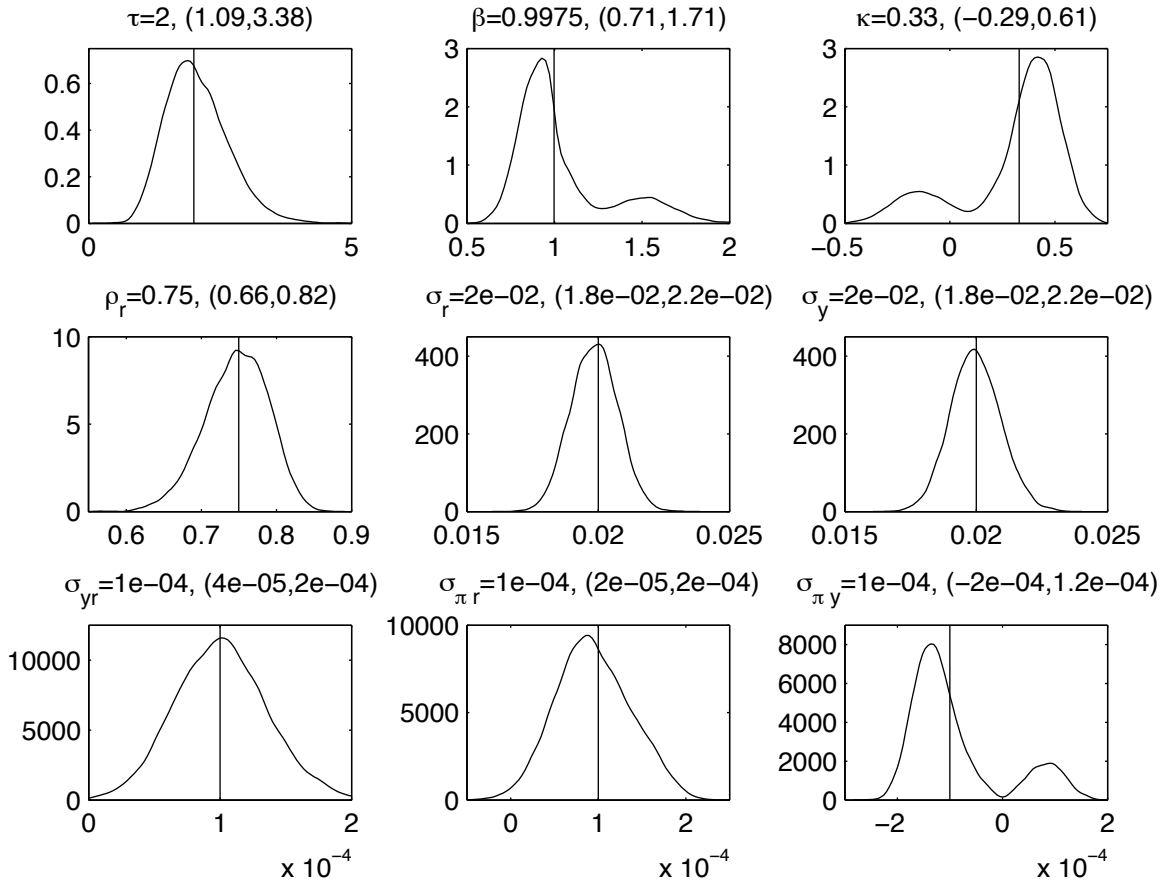


Pseudo code:

---

1. For draw $\widehat{\Pi}_{\mathtt{i}}$, compute both estimators $\widehat{\vartheta}_{\mathtt{i}}$ and $\widehat{\vartheta}^*_{\mathtt{i}}$ by plugging in to g^{-1}.
2. If both are outside of reals, the estimator is said to not exist for the given draw (22% of utilized sample).
3. If both estimators are in the reals
   (a) If both are inside of Θ_0, the analyst can not distinguish between the two, so plot both in Monte Carlo with equal weight 1.
   (b) If both are outside of Θ_0, the analyst can not distinguish between the two, so plot both in Monte Carlo with equal weight 1.
   (c) If one is in Θ_0 and one is outside, plot only the one inside, with weight 2.
4. If one is in the reals and one is outside, plot only the one inside with weight of 2 (Note, this option has no effect in the given sample).

Figure 7: Monte Carlo distribution of Simplified Model MLE with $\beta \approx 0.9975$ identifying restriction at ϑ_0: T=250, N=10,000. Two-sided $(\alpha/2, 1-\alpha/2)$ confidence interval, $\alpha = 5\%$.



Pseudo code:

---

1. For draw $\widehat{\Pi}_{\mathtt{i}}$, compute both estimators $\widehat{\vartheta}_{\mathtt{i}}$ and $\widehat{\vartheta}^*_{\mathtt{i}}$ by plugging in to g^{-1}.
2. If both are outside of reals, the estimator is said to not exist for the given draw (22% of utilized sample).
3. If both estimators are in the reals
   (a) If both imply $\widehat{\beta}_{\mathtt{i}} > 1$, the analyst can not distinguish between the two, so plot both in Monte Carlo with equal weight 1.
   (b) If only one implies $\widehat{\beta}_{\mathtt{i}} > 1$ plot the other in Monte Carlo with weight 2.
   (c) If both imply $\widehat{\beta}_{\mathtt{i}} < 1$
      i. If one estimator yields a $\widehat{\beta}_{\mathtt{i}}$ which is closer to 0.9975, plot only that estimator, with weight 2.
      ii. If both estimators yield $\widehat{\beta}_{\mathtt{i}}$ which are equidistant from 0.9975, plot both estimators, with weight 1 each.
4. If one is in the reals and one is outside, plot only the one inside with weight of 2 (Note, this option has no effect in the given sample).

	θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Π		τ	β	ν	ϕ	Π	ψ_π	ψ_y	ρ_z	ρ_g	ρ_r	σ_z	σ_g	σ_r	σ_{gz}	σ_{rz}	σ_{rg}
1	ϕ_{rr}	×	×	×	×	×	×	×	×		×						
2	ϕ_{yr}	×	×	×	×	×	×	×	×	×	×						
3	$\phi_{\pi r}$	×	×	×	×	×	×	×	×		×						
4	ρ_g									×							
5	$\phi_{r\pi}$	×	×	×	×	×	×	×	×		×						
6	$\phi_{y\pi}$	×	×	×	×	×	×	×	×	×	×						
7	$\phi_{\pi\pi}$	×	×	×	×	×	×	×	×		×						
8	ω_r	×	×	×	×	×	×	×	×		×	×		×		×	
9	ω_y	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×
10	ω_π	×	×	×	×	×	×	×	×		×	×		×		×	
11	ω_{yr}	×	×	×	×	×	×	×	×		×	×		×	×	×	×
12	$\omega_{\pi r}$	×	×	×	×	×	×	×	×		×	×		×		×	
13	$\omega_{\pi y}$	×	×	×	×	×	×	×	×		×	×		×	×	×	×

Table 8: An and Schorfheide (2007) model functional mapping $g : \theta \to \Pi$.

5.4 Full Model

Whereas the mapping from θ to Π in the simplified model was complicated, the corresponding mapping for the L model is simply indecipherable from a human perspective; in the computer code for this paper, mathematical expressions for Π in terms of θ are so complicated, they take up pages of text. However, the expressions are closed-form and calculated by the computer, and therefore guaranteed to be correct. A graphical summary is given by Table 8.

As was the case for the simplified model, it is immediately obvious by the necessary order condition $n_\Pi \geq n_\theta$ that the entire vector θ may not be identified. Since there are only $n_\Pi = 13$ reduced form parameters, but $n_\theta = 16$ structural parameters, at least 3 parameters must be set to constants for even a fighting chance that the remaining 13 are conditionally locally or globally identified. Figuring out which are candidates to be set can be done by using the location of ×'s in the table, and sequential elimination. The parameters that will be set here are exactly those chosen by Komunjer and Ng (2011) in one of their conditional identification schemes, $\overline{\alpha} = (\overline{\nu}, \overline{\phi}, \overline{\psi}_y)$. These are trained to their values in θ_0. The remaining parameters which may be identified are

collected in the 13×1 vector structural parameter

$$\underset{(13\times1)}{\vartheta} = (\tau, \beta, \Pi, \psi_\pi, \rho_z, \rho_g, \rho_r, \sigma_z, \sigma_g, \sigma_r, \sigma_{gz}, \sigma_{rz}, \sigma_{rg})'$$

Since $n_\vartheta = 13 = n_\Pi$, if ϑ is identified at a point, it is exactly identified. Indeed, the 13×13 Jacobian $J(\vartheta_0; \overline{\alpha}) = \partial\Pi(\vartheta; \overline{\alpha})/\partial\vartheta'|_{\vartheta=\vartheta_0}$, is full column rank, thus satisfying a necessary and sufficient condition for ϑ to be locally identified at ϑ_0; that the demeaned model is conditionally identified using $\overline{\alpha}$ was previously shown by Komunjer and Ng (2011). But since ϑ is exactly locally identified, and the solution of the model is analytically derived, one can once again hope to derive g^{-1}. Thus, one can observe directly whether g is injective at ϑ_0. The results for one point are given in Figure 8, and for many points in.

As in the simplified model, even when squared variables such as standard deviations and the steady state of inflation are normalized to positive numbers, there are exactly two ϑ that satisfy $\vartheta = g^{-1}(\Pi; \alpha)$ for any realization of Π. Specifically, while the 7 parameters β, ψ_π, ρ_g, ρ_r, σ_g, σ_r, and σ_{rg} are the same in both solutions, all of the 6 parameters τ, Π, ρ_z, σ_z, σ_{gz}, and σ_{rz} differ. But unlike the analysis of the simple model, this result does not actually in itself imply that ϑ is not globally identified in $\mathbb{R}^{13}$ at ϑ_0. In fact, as listed in Figure 8, the only realization of ϑ which is observationally equivalent to ϑ_0 is $\vartheta_1 \in \mathbb{C}^{13}\backslash\mathbb{R}^{13}$; in row 3, the value for the steady state of inflation Π at ϑ_1 is $4.8i$.[18] So, even though ϑ is not globally identified in all $\Theta \subset \mathbb{C}^{13}$ at ϑ_0, it is globally identified in the strictly smaller, and more economically meaningful space of all $\Theta \subset \mathbb{R}^{13} \subset \mathbb{C}^{13}$. Impulse responses for ϑ_0 and ϑ_0^* are also given in Figure 8, demonstrating that only y_t's response to ε_{gt} is the same for both points. However, while result that ϑ is in fact globally identified in all $\Theta \subset \mathbb{R}^{13}$ at ϑ_0 using $\overline{\alpha}$ is encouraging, this result is specific of the realization ϑ_0, so it is important to assess other points in Θ_0.

Another important point is ϑ_1, listed in Figure 9. Although the impulse responses for ϑ_1 and ϑ_1^* appear to be the same at first glance, all but the response of y_t to ε_{gt} in

[18] For intuition of this result, recall that the Phillips curve coefficient is written $\kappa = \tau(1-\nu)/(\nu\phi\Pi^2)$. Thus, Π is complex simply means that κ is positive; at the same time $0 < \nu < 1$, $\tau < 0$, and $\phi > 0$.

Figure 8: Full model without means observational equivalence in $\mathbb{C}^{13}$ but not Θ_0 (see row 3), and impulse-responses: ϑ_0 ($\circ$) vs ϑ_0^* ($+$). $\Pi = g(x;\overline{\alpha})$; $x = \vartheta_0$ or ϑ_0^*. Missing impulse-responses correspond to exclusion restrictions in $D(\theta)$, see ABCD representation.

		$\boldsymbol{g(x,\overline{\alpha})}$		$\boldsymbol{\vartheta_0}$	$\boldsymbol{\vartheta_0^*}$
1	ϕ_{rr}	0.79	$\boldsymbol{\tau}$	2	**-45.4**
2	ϕ_{yr}	0.19	β	0.9975	0.9975
3	$\phi_{\pi r}$	0.12	$\boldsymbol{\Pi}$	1.008	$\boldsymbol{4.8i}$
4	ρ_g	0.95	ψ_π	1.5	1.5
5	$\phi_{r\pi}$	0.25	$\boldsymbol{\rho_z}$	0.9	**0.51**
6	$\phi_{y\pi}$	-0.46	ρ_g	0.95	0.95
7	$\phi_{\pi\pi}$	0.62	ρ_r	0.75	0.75
8	ω_r	2e-2	$\boldsymbol{\sigma}_z$	3e-2	**1.85**
9	ω_y	5e-4	σ_g	6e-2	6e-2
10	ω_π	7e-4	σ_r	2e-2	2e-2
11	ω_{yr}	8e-2	$\boldsymbol{\sigma_{gz}}$	1e-4	**-1e-2**
12	$\omega_{\pi r}$	2e-3	$\boldsymbol{\sigma_{rz}}$	1e-4	**1e-2**
13	$\omega_{\pi y}$	4e-2	σ_{rg}	-1e-4	-1e-4

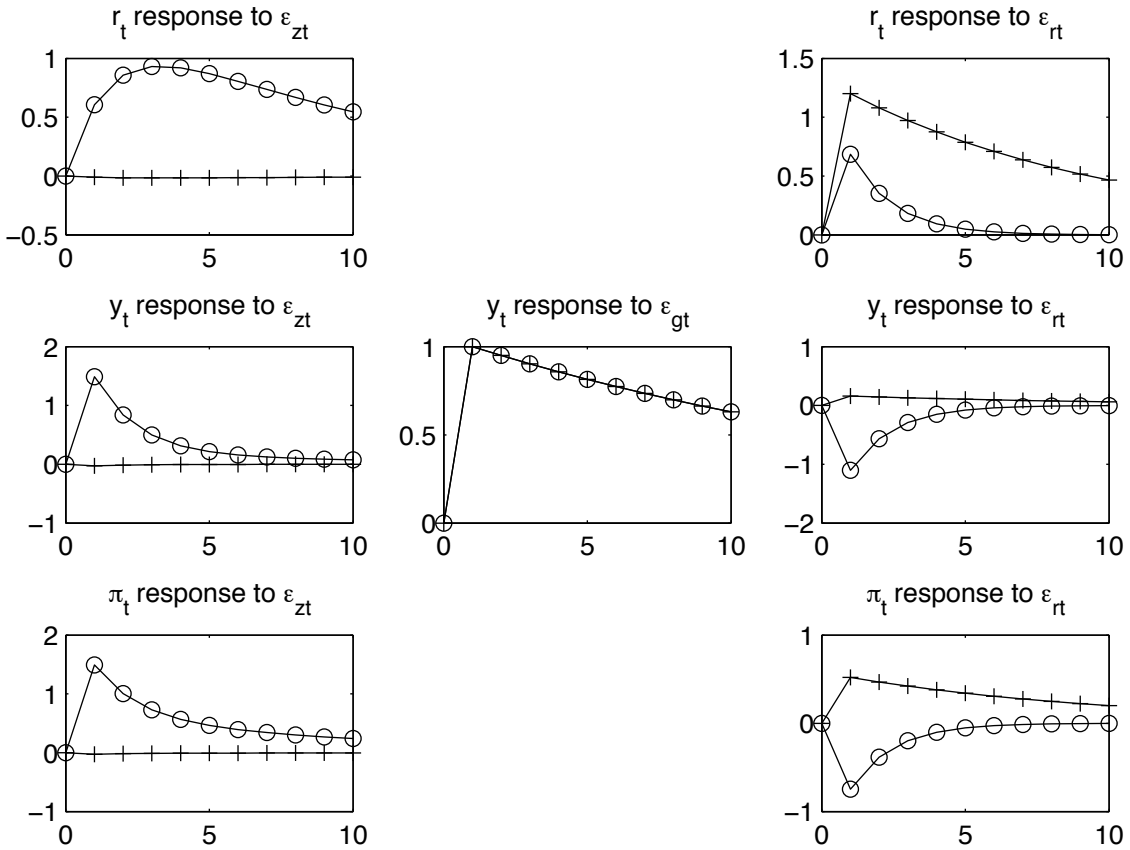
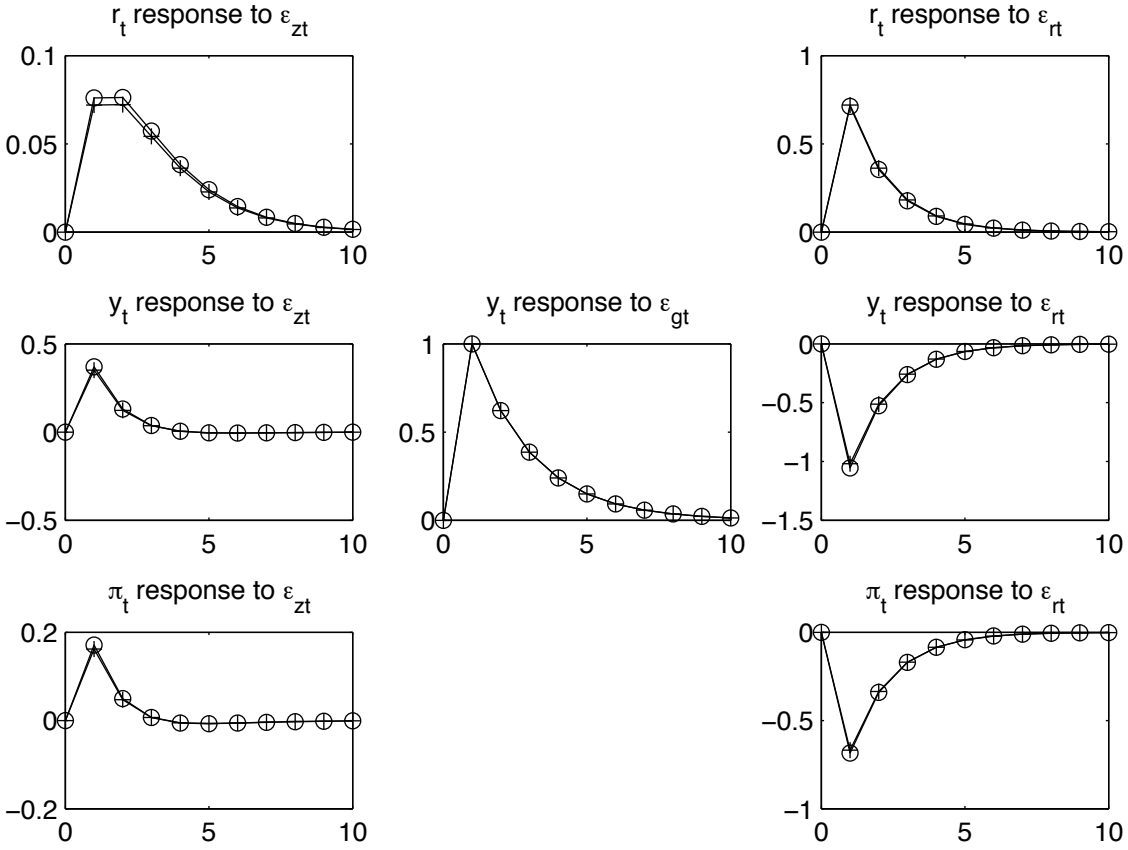


Figure 9: Full model without means observational equivalence in in Θ_0, and impulse-responses: ϑ_1 ($\circ$) vs ϑ_1^* ($+$). $\Pi = g(x;\overline{\alpha})$; $x = \vartheta_1$ or ϑ_1^*.

		$\boldsymbol{g(x,\overline{\alpha})}$		$\boldsymbol{\vartheta_1}$	$\boldsymbol{\vartheta_1^*}$
1	ϕ_{rr}	0.65	$\boldsymbol{\tau}$	1.99	**2.09**
2	ϕ_{yr}	-0.26	β	0.9839	0.9839
3	$\phi_{\pi r}$	-0.14	$\boldsymbol{\Pi}$	1.004	**1.029**
4	ρ_g	0.62	ψ_π	1.21	1.21
5	$\phi_{r\pi}$	0.16	$\boldsymbol{\rho_z}$	0.504	**0.499**
6	$\phi_{y\pi}$	-0.45	ρ_g	0.62	0.62
7	$\phi_{\pi\pi}$	0.35	ρ_r	0.70	0.70
8	ω_r	0.42	$\boldsymbol{\sigma}_z$	0.16	**0.12**
9	ω_y	0.97	σ_g	0.45	0.45
10	ω_π	0.37	σ_r	0.57	0.57
11	ω_{yr}	-0.39	$\boldsymbol{\sigma_{gz}}$	-0.05	**-0.03**
12	$\omega_{\pi r}$	-0.16	$\boldsymbol{\sigma_{rz}}$	0.09	**0.06**
13	$\omega_{\pi y}$	0.35	σ_{rg}	-0.22	-0.22



	Yes	No
Locally Identified	100 %	0%
Globally Identified in Θ_0	99.5%	0.5 %
Globally Identified in All $\Theta \subset \mathbb{R}^{13}$	59.4%	40.6 %

Table 9: Identification of ϑ in full model without means: 100,000 points drawn from Θ_0.

fact differ from point-to-point by very small amounts (all crosshairs are not centered; refer also to the impulse responses in Figure 8 for clarity of this point). Thus, even though both points imply similar economic implications, any determination between the two is not technically a normalization as defined in Definition 4, but something stronger. Furthermore, notice that ϑ_1 and ϑ_1^* are both contained in Θ_0. Therefore, these bounds alone may not be used as an identifying restriction as they may be to distinguish between ϑ_0 and ϑ_0^*. An identifying restriction which would work is to simply select the estimator that yields a τ closer to 2, in this case, ϑ_1. Much as the β criterion for the O model reflected tight priors on 0.9975, this similarly reflects the common tight prior on $\tau = 2$.

As was the case for the simplified model, it is of interest to understand how many points are like ϑ_0 versus ϑ_1 in the sense of which identification scheme is successful. Local and global identification is assessed on a uniform grid of Θ in Table 9. Once again, ϑ is locally identified for all draws, but now is also globally identified in Θ_0 for nearly all draws. In addition, ϑ is globally identified in all of $\mathbb{R}^{13}$ in roughly 6/10 draws. So, even though ϑ_0 is part of this majority, there is evidently a large portion of draws at which ϑ is not globally identified in all reals. Yet, again, the parameter space Θ_0 may be used as a reasonable basis to eliminate many of the 40% of points not identified in all of the reals. In fact, only 0.05% of draws yield an observationally equivalent point also in Θ_0. In other words, points like ϑ_1 exist, but are relatively rare compared to points like ϑ_0.

Finally, Monte Carlo distributions are computed. The typical MLE is presented in Figure 10 and the $\tau \approx 2$ identification criterion is enforced in Figure 11. In this case, the latter identifying restriction has the added bonus of curtailing a significant fat and skewed tail for $\widehat{\tau}$ observed in Figure 10. Also, note that the distribution under the

Θ_0-based identification scheme, corresponding to Figure 6 for the simplified model, is exactly the same as Figure 10, since there are zero instances in which one estimator is in Θ_0 and one is out. Finally, notice that in all Figures 10-11 there are substantial biases in the estimators also documented on a smaller scale previously for the simplified model. Since the estimator I have utilized is based on plugging in OLS estimates, there is no risk that this is a numerical quirk, but rather a reflection of small sample bias. One particularly prominent feature is that the mode of $\widehat{\beta}$ is greater than 1. Thus, for identification schemes like the $\beta \approx 1$ approach used in the simplified model to work, it is essential to pair estimation with bias correction. I consider this momentarily.

6 Conclusion

In this paper I have developed entirely new, simple, reliable, and computationally efficient tools for empirical analysis of a specific well-known DSGE model. As I have shown, global identification failures are eminent, but suitable normalizations based on macroeconomic priors are typically available. Although these normalizations may cause the distribution of the maximum likelihood estimator to be nonstandard, the tools I provide also allow the analyst to compute the small sample distribution of the MLE under these restrictions. While discovering global identification failures in DSGE models is difficult in itself, this paper has also argued that it is of equal importance to account for the statistical implications of a given identification procedure.

Figure 10: Distribution of L model without means MLE at ϑ_0: T=250, N=10,000. Two-sided $(\alpha/2,1-\alpha/2)$ confidence interval, $\alpha = 5\%$.

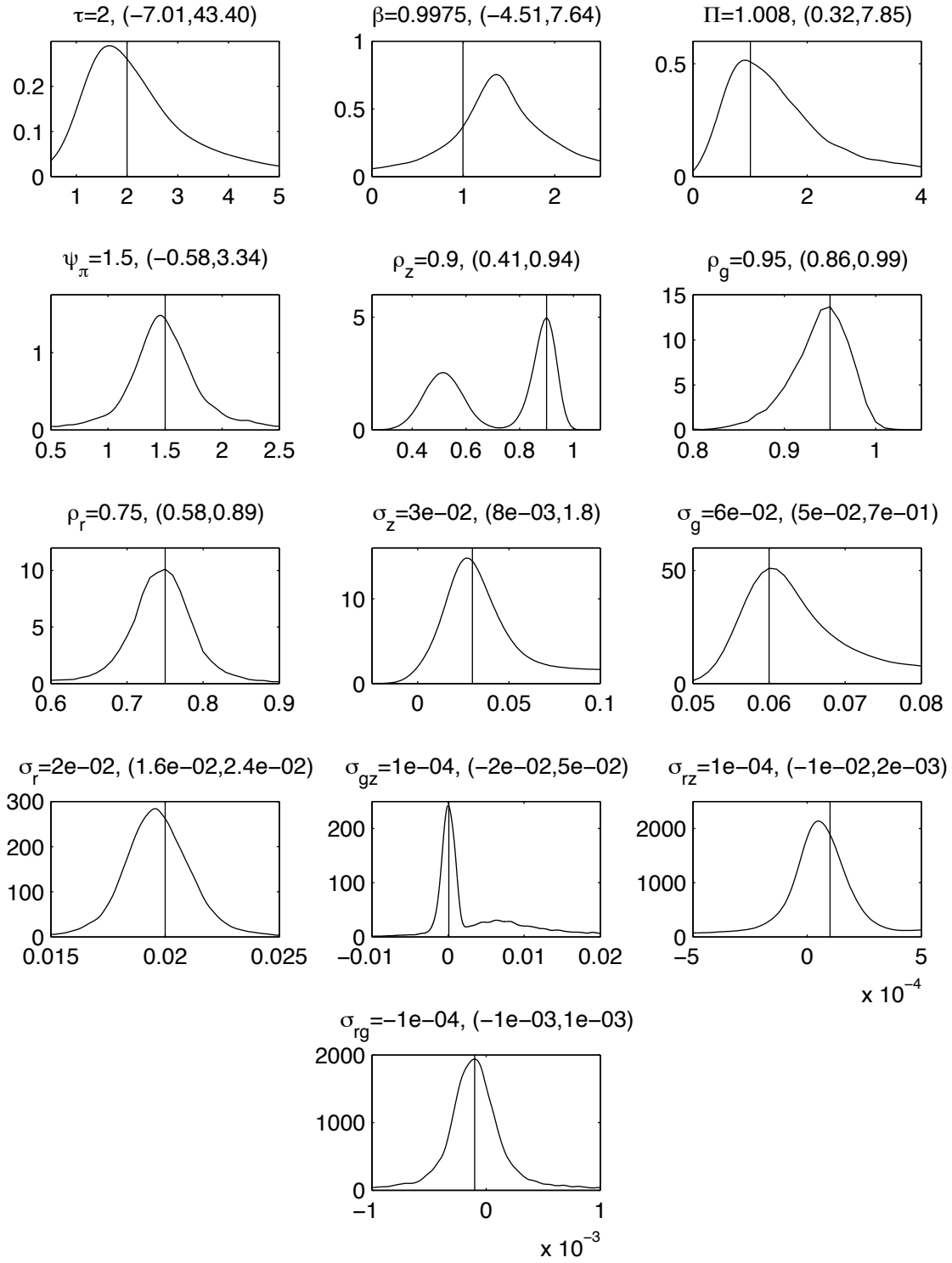
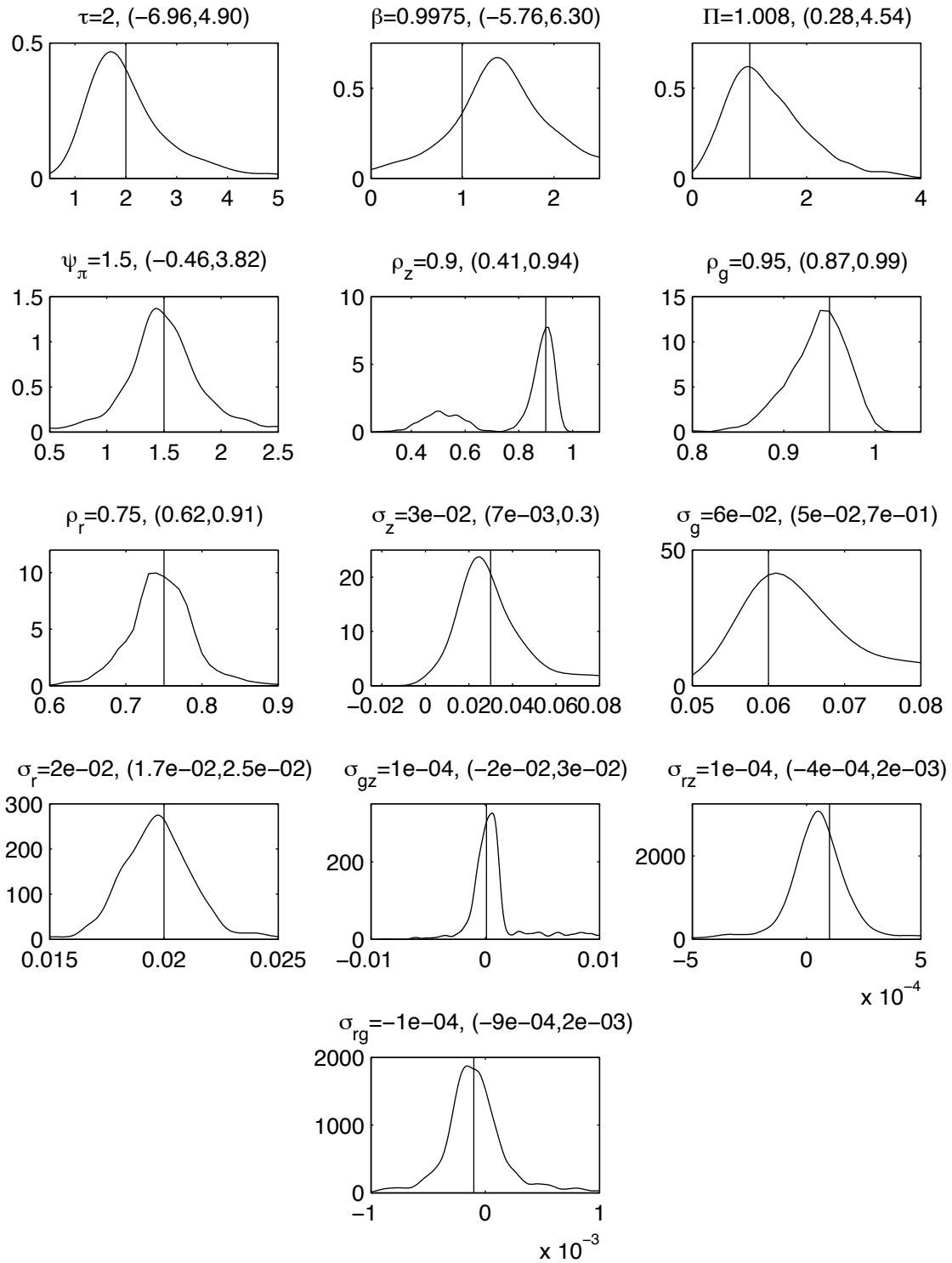


Figure 11: Distribution of L model without means MLE with $\tau \approx 2$ identification scheme ϑ_0: T=250, N=10,000. Two-sided $(\alpha/2, 1-\alpha/2)$ confidence interval, $\alpha = 5\%$.



References

Abadir, K. M. and J. R. Magnus (2005): *Matrix Algebra*, Cambridge University Press.

An, S. and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26, 113–172.

Andreasen, M. M. (2010): "How to Maximize the Likelihood Function for a DSGE Model," *Computational Economics*, 35, 127–54.

Andrews, I. and A. Mikusheva (2013): "Maximum Likelihood Inference in Weakly Identified DSGE Models," *Under Review, Quantitative Economics*.

Beyer, A. and R. E. Farmer (2006): "Identification Problems in SDGE Models with an Illustration to a Small Macro Model," *Computing in Economics and Finance 2006*, 81.

Bianchi, F. (2013): "Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics," *Review of Economic Statistics*, 80, 463–90.

Canova, F. and L. Sala (2009): "Back To Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431–449.

Chari, V., P. J. Kehoe, and E. R. McGrattan (2009): "New Keynesian Models: Not Yet Useful for Policy Analysis," *American Economic Journal: Macroeconomics*, 1, 242–266.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

Christiano, L. J., M. Trabandt, and K. Walentin (2010): "DSGE Models for Monetary Policy Analysis," *NBER Working Paper No. 16074*.

Cochrane, J. (2011): "Determinacy and Identification with Taylor Rules," *Journal of Political Economy*, 119, 565–615.

Doh, T. (2011): "Yield Curve in an Estimated Nonlinear Macro Model," *Journal of Economic Dynamics and Control*, 35, 1229–1244.

Dufour, J.-M., L. Khalaf, and M. Kichian (2013): "Identification-Robust Analysis of DSGE and Structural Macroeconomic Models," *Journal of Monetary Economics*, 60, 340–50.

Fernández-Villaverde, J., J. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021–1026.

Fisher, F. (1966): *The Identification Problem in Econometrics*, McGraw-Hill.

Fukač, M., D. F. Waggoner, and T. Zha (2007): "Local and Global Identification of DSGE Models: A Simultaneous-Equation Approach," *Working Paper*.

Gale, D. and H. Nikaidô (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159, 81–93.

Galí, J. (2008): *Monetary Policy, Inflation, and the Business Cycle*, Princeton University Press.

Galí, J., F. Smets, and R. Wouters (2011): "Unemployment in and Estiamted New Keynesian Model," in *Unemployment in an Estimated New Keynesian Model*, NBER Macroeconomics Annual, 329–60.

Giacomini, R. (2013): "The Relationship Between DSGE and VAR Models," *Advances in Econometrics*, 31.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 4, 197–229.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Hamilton, J. D., D. Waggoner, and T. Zha (2007): "Normalization in Econometrics," *Econometric Reviews*, 26, 221–252.

Hamilton, J. D. and C. Wu (2012): "Identification and Estimation of Gaussian Affine Term Structure Models," *Journal of Econometrics*, 168, 315–31.

Ireland, P. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, 28, 1205–1226.

Iskrev, N. (2010): "Local Identification in DSGE Models," *Journal of Monetary Economics*, 57, 189–202.

Kailath, T., A. H. Sayed, and B. Hassibi (2000): *Linear Estimation*, Prentice Hall.

Kleibergen, F. and S. Mavroeidis (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Economic and Business Statistics*, 27, 293–311.

Kocięcki, A. and M. Kolasa (2013): "Global Identificaiton of Linearized DSGE Models," *Working Paper*.

Komunjer, I. and S. Ng (2011): "Dynamic Identification of DSGE Models," *Econometrica*, 79, 1995–2032.

Koopmans, T. C. and O. Reiersøl (1950): "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21.

Kydland, F. E. and E. C. Prescott (1982): "Time to Build and Aggregate Fluctutations," *Econometrica*, 50, 1345–1370.

McCallum, B. T. (1983): "On Non-Uniqueness in Rational Expectations Models," *Journal of Monetary Economics*, 11, 139–168.

——— (1999): "Roles of the Minimal State Variable Criterion in Rational Expectations Models," *NBER Working Paper Series*, 7087.

Nelson, C. and R. Startz (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967–76.

Qu, Z. (2013): "Inference in DSGE Models with Possible Weak Identification," *Quantitative Economics*, Forthcoming.

Qu, Z. and D. Tkachenko (2012): "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 3, 95–132.

——— (2013): "Local and Global Parameter Identification in DSGE Models Allowing for Indeterminacy," *Working Paper*.

Rasche, R. H. and H. T. Shapiro (1968): "The F.R.B.-M.I.T. Econometric Model: Its Special Features," *American Economic Review*, 58, 123–49.

Ravenna, F. (2007): "Vector Autoregressions and Reduced Form Representations of DSGE Models," *Journal of Monetary Economics*, 54, 2048–2064.

Ríos-Rull, J.-V., F. Schorfheide, C. Fuentes-Albero, M. Kryshko, and R. Santaeulàlia-Llopis (2012): "Methods Versus Substance: Measuring the Effects of Technology Shocks on Hours," *Journal of Monetary Economics*, 59, 826–846.

Rothenberg, T. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–91.

——— (1973): *Efficient Estimation With A Priori Information*, Yale University Press.

Rudebusch, G. D. and E. T. Swanson (2012): "The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks," *American Economic Journal: Macroeconomics*, 4, 105–43.

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

——— (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

Smets, F. and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–75.

Stock, J. H., M. W. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.

van Binsbergen, J. H., J. Férnanedez-Villaverde, R. S. J. Koijen, and J. F. Rubio-Ramírez (2012): "The Term Structure of Interest Rates in a DSGE Model with Recursive Preferences," *Journal of Monetary Economics*, 59.

Wald, A. (1950): "Note on the Identification of Economic Relations," in *Statistical Inference in Dynamic Economic Models*, New York: John Wiley.

Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press.