Maximum likelihood estimation of globally unidentified dynamic stochastic general equilibrium models using a-priori information*

Stephen D. Morris†
Department of Economics
Bowdoin College

December 9, 2014

Abstract

The issue of global identifiability arises when one may not rule out observational equivalence in the entirety of a parameter space. In this paper, I prove analytically that popularly utilized restrictions for local identifiability in DSGE models do not imply global identifiability, and that the policy implications of observationally equivalent points are distinct. Furthermore, this gives rise to non-normality in the small sample distribution of the MLE which might otherwise be miscategorized as evidence of weak identifiability. I then offer a pragmatic solution: Priors from macroeconomic theory partition observationally equivalent points by likeliness, and may be used to construct a modest generalization to the MLE which both has good small sample properties, and is consistent and efficient, without global identifiability. The estimator is easy to implement – in fact, unknowingly already in use – and the MLE obtains in the null.

JEL Classifications: C32, C51, E12, E52.

Keywords: DSGE, Maximum Likelihood Estimation, Identification.

*I would foremost like to thank my dissertation advisor at UC San Diego, James D. Hamilton. I also benefited from the support of the remaining members of my committee, Thomas Barranga, Ivana Komunjer, Valerie Ramey, and Ross Valkanov, as well as feedback from seminar participants at UC San Diego, Bowdoin College, and the 2014 Annual Meeting of the Society for Economic Dynamics at the University of Toronto. A previous version of this paper circulated as Chapter 1 of the author's dissertation, "Global identification in DSGE models." MATLAB documentation to replicate all computational claims is available on the author's website.

†Bowdoin College, Department of Economics, 9700 College Station, Brunswick, ME 04011. smorris@bowdoin.edu.

1 Introduction

Dynamic stochastic general equilibrium (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 maximization 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 also been called into question.[4]

In response to this concern, two related literatures have emerged. The first, including Iskrev (2010), Komunjer and Ng (2011), and Qu and Tkachenko (2012), considers local identifiability of DSGE parameters. A parameter point θ is not locally identifiable when it is not uniquely distinguishable in an ϵ-neighborhood of θ. This set of papers has therefore considered, for example, which parameters must be set to constants to achieve conditional local identifiability of the complement subset. The second literature, including Dufour et al. (2013), Guerron-Quintana et al. (2013), Andrews and Mikusheva (2014), and Qu (2014), considers weak identification. Weak identification causes statistics to be non-normally distributed even in very large samples (Bound et al., 1995). This set of papers has therefore provided a number of robust inferential procedures. Taken together, these two literatures have afforded new rigor to empirical analysis of DSGE models.

[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 December 2014, the representative paper of Smets and Wouters (2003) had over 3000 scholarly citations on Google Scholar.

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

[3] Early examples include Smets and Wouters (2003) and Ireland (2004).

[4] Kleibergen and Mavroeidis (2009) consider weak identification of coefficients in the Phillips curve, Cochrane (2011) considers non-identification of coefficients of the Taylor rule, and Beyer and Farmer (2006) and Canova and Sala (2009) consider several types of identifiability in 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 necessity for any argument for or against empirical efficacy.

Yet, as critical as identifiability itself is to take stock in the intended limits of these results. For instance, local identifiability is neither sufficient for consistency of the maximum likelihood estimator, nor consquently to rule out contradictory policy implications on the basis of the data. In this paper, first, I clarify that restrictions popularly utilized to obtain local identifiability in DSGE models do not generally imply global identifiability. Specifically, I will consider the de facto typical contemporary analyst who, armed with existing rank and order conditions, reasonably fixes as few parameters to constants as is necessary to yield conditional local identifiability of the complement subset. As I prove analytically, such restrictions only necessarily provide set-identification over multiple points not sharing a locality. Furthermore, these points yield distinct macroeconomic consequences, in the form of monetary policy impulse-responses: While one point of interest implies that increasing interest rates is contractionary, its observationally equivalent point paradoxically suggests expansion. Thus, even the very elemental purposes of DSGE models become suspect. Furthermore, in this case, the MLE is multiplicatively-valued, and therefore trivially not consistent.[5]

In addition, I show that the subsequent small sample distribution of the MLE is markedly non-normal. In particular, the distributional profile is typically multimodal. This result is in some sense expected, given consistency is a formal requirement for asymptotic normality. However, it is an instructive finding because non-normality in small samples is sometimes interpreted as evidence of weak identification, specifically. Such evidence-based conclusions are perfectly reasonable, given the results of Nelson and Startz (1990). They demonstrated dramatic bimodality of the instrumental variables estimator with irrelevant instruments. Yet, there exist no formal tests for weak identification in DSGE models.[6] Inasmuch, these findings illustrate that it is impossible to conclude weak identification is the unique cause of observed small sample non-normality, or other normatively poor behavior of the MLE. One must also consider the primitive assumption of global identifiability; when the parameters of the same models I consider are globally identifiable, the small sample distribution does not signficantly depart from normal.

[5]Consquently, there exists no formal theory from which to interpret any point statistic from a Bayesian posterior – such as the mean or mode – as asymptotically equivalent to a consistent estimator.

[6]Informal approaches to assessing strength of identification include numerically evaluating the slope of the likelihood (Canova and Sala, 2009), computing the concentration parameter (Kleibergen and Mavroeidis, 2009) and, extrapolating on the work of Qu (2014), observing the magnitude of eigenvalues of the information matrix as the sample size increases.

Finally, I offer a pragmatic solution to the problem of maximum likelihood estimation of DSGE models when parameters are possibly not globally identifiable. I introduce an intuitive and easy-to-implement generalization to the MLE which is consistent, asymptotically normal, and efficient, without the need for global identifiability. The only requirement in place of identifiability is that the analyst be able to specify priors which rank the population value of the parameters as a-priori more likely than its observationally equivalent points. As I will show, this assumption is general; those independent priors typically used to construct posterior distributions for DSGE models, or sign restrictions for impulse-responses, typically suffice. Importantly, the analyst need not even know the location of nuisance observationally equivalent points. This is essential, since locating all such points in unbounded parameter spaces like the real numbers is usually infeasible, and difficult even in bounded cases. With identifiability – known or unbeknownst to the analyst – the standard MLE is obtained as a special case, independent of whether the priors are correctly specified. The estimator also has good small sample properties, and is easy to implement computationally; in fact, as I will describe, it is already unknowingly sometimes in use.

There are currently three other working papers considering global identification in DSGE models, by Fukač et al. (2007), Kocięcki and Kolasa (2014), and Qu and Tkachenko (2014). These papers are primarily concerned with providing methodologies for discerning global identifiability in DSGE models, which neither have analytical solutions nor reduced form representation generally, and demonstrating these results in examples; the subject of the latter two papers is the model of An and Schorfheide (2007). These results are of critical importance for discerning when local identifiability is, or is not, sufficient for estimation and inference. I am concerned primarily with how the analyst should proceed with estimation without global identifiability, for example, when the results of one of the previously mentioned papers suggests as much. Therefore, the logical progression of this paper is opposite. First, I show that the An and Schorfheide model itself actually does have both an analytical solution and reduced form representation. Then, I use this special case to demonstrate that restrictions for local identifiability do not ensure global. Finally, I generalize the intuition gleaned from this special case to suggest the previously noted generalization to the MLE, which is applicable to all DSGE models for which the likelihood function is defined.

In the next section, I consider analytical solutions and reduced form representation

of DSGE models, beginning with the small-scale model presented in Canova and Sala (2009). I consider points of both determinacy and indeterminacy, which will prove pivotal to subsequent identification analysis.

2 Solutions and reduced form representation

In this section I consider modest generalizations of two linearized and Gaussian models. The first is the simple specification with no latent variables presented in Canova and Sala (2009). The second, from An and Schorfheide (2007), contains latent variables. I choose these models because they are both well-known to the existing literature on DSGE identification. However, in contrast with the assumptions of previous contributions, I will reveal that each model has multiple analytical solutions and restricted Gaussian VAR(1) representation. Thus, each is possibly indeterminate, and presented in reduced form. Specifically, where θ denote the structural parameters of a given model, and $\pi^{(i)}$ denote the reduced form parameters corresponding to solution i, I will for both models calculate analytically the moment condition

$$g(\pi^{(i)}, \theta) = 0 \tag{1}$$

The closed-form function g provides the starting point for global identifiability analysis.

2.1 Canova and Sala (2009) model

Let us consider a simple model not explicitly derived from microfoundations. It consists of a Taylor rule, aggregate demand equation, and Phillips curve.

$$r_t = \rho_r r_{t-1} + \psi_\pi E_t \pi_{t+1} + \varepsilon_{rt} \tag{2}$$

$$y_t = E_t y_{t+1} - (1/\tau)(r_t - E_t \pi_{t+1}) + \varepsilon_{yt} \tag{3}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \varepsilon_{\pi t} \tag{4}$$

r_t is the nominal interest rate, y_t is nominal output, and π_t is inflation. τ is the coefficient of relative risk aversion, β is the discount factor, and κ is a function of the degree of price stickiness. ρ_r and ψ_π are indicative of policy activeness. These 5

structural parameters are collected in

$$\underset{5\times1}{\theta_s} = (\tau, \beta, \kappa, \psi_\pi, \rho_r)' \tag{5}$$

The structural shocks to monetary policy ε_{rt} and technology ε_{yt}, as well as the cost-push shock $\varepsilon_{\pi t}$, are collected in the vector $\varepsilon_t = \begin{bmatrix}\varepsilon_{rt} & \varepsilon_{yt} & \varepsilon_{\pi t}\end{bmatrix}'$, which is distributed $\varepsilon_t \sim NID(0, \Sigma)$ for

$$\text{vech}(\Sigma)' = \underset{6\times1}{\theta_\sigma} = \left(\sigma_r^2, \sigma_{yr}, \sigma_{\pi r}, \sigma_y^2, \sigma_{\pi y}, \sigma_\pi^2\right)' \tag{6}$$

NID denotes iid Gaussian, $\{\sigma_i\}$ are standard deviations, $\{\sigma_{ij}\}$ are covariances. The vech operator selects only the elements of a matrix on and below the diagonal. I consider variances as structural parameters rather than standard deviations to avoid the need for normalization in squaring when shortly considering global identifiability. The total 11 structural parameters of the model are thus

$$\underset{11\times1}{\theta} = \left(\theta_s', \theta_\sigma'\right)' \tag{7}$$

This model is similar to that studied by Canova and Sala (2009) with three exceptions. First, I have allowed lagged interest rates to enter the Taylor rule. Second, I have allowed the disturbances to be correlated. Both are generalizations of the original specification. Third, I have assumed that the data is without means. I have made this assumption only for parsimony of the arguments, and will comment on the role means for identifiability subsequently. Henceforth, I refer to the above as the CS model.

2.1.1 Solution

I now solve the CS model. Let us collect the observable variables in the 3×1 vector $Y_t = \begin{bmatrix}r_t & y_t & \pi_t\end{bmatrix}'$. Since r_t is the only lagged variable, the solution is of the form

$$E_t Y_{t+1} = \phi(\theta_s) r_t \tag{8}$$

$\phi = \begin{bmatrix}\phi_{rr} & \phi_{yr} & \phi_{\pi r}\end{bmatrix}'$, where ϕ_{ij} are scalar functions of θ_s. By the method of undetermined coefficients (See Galí (2008)), (8) with (2)-(4) implies the 3×1 system of

implicit functions $g^{\phi}(\phi, \theta_s) = 0$ with rows

$$g_1^{\phi}(\phi, \tau) = \phi_{yr} - (\phi_{yr} - 1/\tau(1 - \phi_{\pi r}))\,\phi_{rr} = 0 \tag{9}$$

$$g_2^{\phi}(\phi, \beta, \kappa) = \phi_{\pi r} - \beta\phi_{\pi r}\phi_{rr} - \kappa\phi_{yr} = 0 \tag{10}$$

$$g_3^{\phi}(\phi, \psi_{\pi}, \rho_r) = \phi_{rr} - \rho_r/(1 - \psi_{\pi}\phi_{\pi r}) = 0 \tag{11}$$

The handwritten solution to g^{ϕ} for ϕ is deceptively elaborate, and prone to human error. However, it is in fact straightforward to obtain a precise solution using symbolic computation. I use MATLAB's symbolic computation toolbox. Proceeding this way reveals that there are exactly three unique solutions for any θ_s, denoted $\phi^{(1)}$, $\phi^{(2)}$, and $\phi^{(3)}$. I do not replicate the overwrought functional forms of the solutions here.[7] Yet, importantly, these functions are known to the computer in closed form.

Each $\phi^{(i)}$ implies the unique restricted VAR(1) representation

$$Y_t^{(i)} = \begin{bmatrix} \phi^{(i)} & \underset{3\times 2}{0} \end{bmatrix} Y_{t-1}^{(i)} + U_t^{(i)} \tag{12}$$

where $U_t^{(i)} \sim NID(0, \Omega^{(i)}(\theta))$ is a 3×1 vector of reduced form disturbances with

$$\underset{6\times 1}{\text{vech}(\Omega^{(i)})} = \omega^{(i)} = D_3^{+}(\delta^{(i)} \otimes \delta^{(i)})D_3 \times \theta_{\sigma} \tag{13}$$

$$\delta^{(i)}(\phi^{(i)}, \rho_r, \kappa) = \begin{bmatrix} \frac{\phi_{rr}^{(i)}}{\rho_r} & 0 & 0 \\ \frac{\phi_{yr}^{(i)}}{\rho_r} & 1 & 0 \\ \frac{\phi_{\pi r}^{(i)}}{\rho_r} & \kappa & 1 \end{bmatrix} \tag{14}$$

$\otimes$ is the Kronecker product, D_3 is the 3-dimensional duplication matrix, and D_3^{+} is its Moore-Penrose pseudoinverse (See Abadir and Magnus (2005)). Since $U_t^{(i)}$ is Gaussian, $\phi^{(i)}$ and $\omega^{(i)}$ are reduced form parameters.[8] They are henceforth collected in the 9×1 vector

$$\underset{9\times 1}{\pi^{(i)}} = \left(\phi^{(i)\prime}, \omega^{(i)\prime}\right)' \tag{15}$$

[7] Indeed, the functional forms would take up several pages. The interested reader is directed to `uap/run/cs/solve/solns` in the paper's MATLAB documentation available freely online. These files are automatically written from the solution to (9)-(11) by executing `.../solve/solveAnCS.m`, guaranteeing no human error in derivations.

[8] Broadly speaking, parameters are reduced form when they are necessary and sufficient statistics from which to write the likelihood function. I will consider the formal definition and usefulness with respect to identifiability in the following section.

Whereas $\phi^{(i)}$ depends only on θ_s, $\omega^{(i)}$ depends on the entire vector θ. Thus, as g^{ϕ} defined a mapping between θ_s and $\phi^{(i)}$, (13) implies a 6×1 system of implicit functions defining the elements of θ_σ, given θ_s is known:

$$g^{\omega}(\pi^{(i)}, \theta) = \theta_\sigma - D_3^+ \left(\delta^{(i)} \otimes \delta^{(i)}\right)^{-1} D_3 \times \omega^{(i)} = 0 \tag{16}$$

$\delta^{(i)}$ is invertible because it is lower-triangular. The main object of interest from the perspective of identifiability of the structural parameters θ is the mapping from (7) to (15), defined by the system of 9 equations $\begin{bmatrix} g^{\phi'} & g^{\omega'} \end{bmatrix}' = 0$, or more concisely put, Equation (1).

2.1.2 Determinacy and indeterminacy

The solution of a DSGE model is called indeterminate when there are multiple solutions for a given θ which imply stable equilibria. The fact that there are three solutions for any θ does not itself imply indeterminacy for all $\theta \in \Theta \subseteq \mathbb{R}^{11}$. Rather, the stability properties of the VAR(1) representation (12) must be checked on a case-by-case basis. This is easily accomplished. The eigenvalues $\{\lambda\}$ of $\left[\begin{array}{c|c} \phi^{(i)} & 0_{3\times 2} \end{array}\right]$ are defined as the solutions to the characteristic equation

$$(\lambda - \phi_{rr}^{(i)})\lambda^2 = 0 \tag{17}$$

The roots are $\{\phi_{rr}^{(i)}, 0, 0\}$. Therefore, a necessary and sufficient condition for stability is simply that $|\phi_{rr}^{(i)}| < 1$. In Table 1, I evaluate determinacy at two points of interest, θ_0 and θ_1, on this basis. Under scenario θ_0, the Taylor rule reaction to expected inflation is relatively hawkish, $\psi_\pi = 1.5$. Therefore, this parameterization is called "active" monetary policy. Under the opposing scenario θ_1, all values are the same beside $\psi_\pi = 0.1$. This parameterization is therefore called "passive" monetary policy. These opposing monetary policy regimes are interesting with respect to one another, because Lubik and Schorfheide (2004) have previously shown that passive monetary policy may lead to indeterminacy. I verify this finding in this model as well. Of the three possible solutions, only one is stable at θ_0, implying determinacy. However, two are stable at θ_1, implying indeterminacy. Therefore, this model is determinate at some points, but indeterminate at others. Finally, Table 1 also serves to prove the analytical

solutions are correct. At θ_0, Solution 1 is exactly that provided by Sims (2002)'s `gensys` routine. At θ_1, Sims's routine returns the error message, "`Indeterminacy. 1 loose endog errors`."

In this section, I have shown that the CS model has restricted VAR(1) representation, and multiple analytical solutions. While it is unsurprising that models with no latent variables such as CS have analytical solutions and reduced form representation, less well-appreciated is the fact that models with latent variables may as well.

2.2 An and Schorfheide (2007) model

The linearized equilibrium conditions of the An and Schorfheide's well-studied model are given by the following six equations. A complete derivation appears in Appendix B. Variable and parameter names are standard, and given there in Table B.1; for example, τ is the CRRA and β is the discount factor. A distinction is shocks to total factor productivity, ε_{zt}, and government spending, ε_{gt}.

$$
\begin{aligned}
z_t &= \rho_z z_{t-1} + \varepsilon_{zt} && (18)\\
g_t &= \rho_g g_{t-1} + \varepsilon_{gt} && (19)\\
r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_\pi \pi_t + (1-\rho_r)\psi_y (y_t - g_t) + \varepsilon_{rt} && (20)\\
y_t &= E_t y_{t+1} + g_t - E_t g_{t+1} - (1/\tau)(r_t - E_t \pi_{t+1} - E_t z_{t+1}) && (21)\\
\pi_t &= \beta E_t \pi_{t+1} + \kappa (y_t - g_t) && (22)\\
c_t &= y_t - g_t && (23)
\end{aligned}
$$

The total 8 structural parameters defining Equations (18)-(23) are[9]

$$\underset{8\times 1}{\theta_s} = (\rho_g, \beta, \kappa, \tau, \rho_z, \psi_y, \psi_\pi, \rho_r)' \tag{24}$$

The vector of shocks $\varepsilon_t = \begin{bmatrix} \varepsilon_{zt} & \varepsilon_{gt} & \varepsilon_{rt} \end{bmatrix}'$ is distributed $NID(0, \Sigma)$ for

$$\text{vech}(\Sigma)' = \underset{6\times 1}{\theta_\sigma} = \left(\sigma_z^2, \sigma_{gz}, \sigma_g^2, \sigma_{rz}, \sigma_{rg}, \sigma_r^2\right)' \tag{25}$$

[9]The microfoundations imply $\kappa = \tau(1-\nu)/(\nu\phi\Pi^2)$ for ν the inverse elasticity of demand, ϕ the index of price stickiness, and Π the steady state of inflation. ν, ϕ, and Π are not separately identifiable in this case. Therefore, here I consider the situation where κ is itself a structural parameter. I consider the separate identifiability of the elements of κ using the means of the data in Appendix B.2, as will be described.

Determinacy: Active monetary policy.

θ_0 :

τ	β	κ	ψ_π	ρ_r	$100\sigma_r$	$100\sigma_y$	$100\sigma_\pi$	ρ^σ_{yr}	$\rho^\sigma_{\pi r}$	$\rho^\sigma_{\pi y}$
2	0.9975	0.33	1.5	0.75	0.2	0.2	0.2	0.5	0.5	-0.5

Solution i	$\begin{bmatrix}\phi^{(i)} & 0\end{bmatrix}$	1e5 $\times\ \Omega^{(i)}$	Stable?
1	$\begin{bmatrix} 0.47 & 0 & 0 \\ -0.62 & 0 & 0 \\ -0.39 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.16 & \cdot & \cdot \\ -0.08 & 0.34 & \cdot \\ 0.04 & -0.22 & 0.14 \end{bmatrix}$	yes
2	$\begin{bmatrix} 1.10+0.6i & 0 & 0 \\ 0.27-0.7i & 0 & 0 \\ 0.32+0.2i & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.13 & \cdot & \cdot \\ 0.22-0.8i & 0.9 & \cdot \\ 0.73-0.2i & 0.08+0.5i & 0.64 \end{bmatrix}$	no
3	$\begin{bmatrix} 1.10-0.6i & 0 & 0 \\ 0.27+0.7i & 0 & 0 \\ 0.32-0.2i & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.13 & \cdot & \cdot \\ 0.22+0.8i & 0.90 & \cdot \\ 0.73+0.2i & 0.08-0.5i & 0.64 \end{bmatrix}$	no

Indeterminacy: Passive monetary policy.

θ_1 : Same as θ_0, but with $\psi_\pi = 0.1$.

Solution i	$\begin{bmatrix}\phi^{(i)} & 0\end{bmatrix}$	1e5 $\times\ \Omega^{(i)}$	Stable?
1	$\begin{bmatrix} 0.85 & 0 & 0 \\ 0.54 & 0 & 0 \\ 1.19 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.52 & \cdot & \cdot \\ 0.56 & 0.9 & \cdot \\ 1.02 & 0.90 & 2.16 \end{bmatrix}$	yes
2	$\begin{bmatrix} 0.62 & 0 & 0 \\ -2.55 & 0 & 0 \\ -2.18 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.27 & \cdot & \cdot \\ -0.95 & 3.66 & \cdot \\ -0.74 & 2.4 & 2.14 \end{bmatrix}$	yes
3	$\begin{bmatrix} 1.43 & 0 & 0 \\ -6.22 & 0 & 0 \\ 4.77 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.46 & \cdot & \cdot \\ -5.97 & 24.65 & \cdot \\ 5.37 & -22.13 & 19.88 \end{bmatrix}$	no

Table 1: Determinacy and indeterminacy in CS. ρ^σ_{ij} denotes the correlation $\sigma_{ij}/(\sigma_i\sigma_j)$.

Thus, the total set of structural parameters is

$$\underset{14\times 1}{\theta} = \left(\theta_s', \theta_\sigma'\right)' \tag{26}$$

Again, I have generalized the model by allowing correlations between shocks to be nonzero. Henceforth, I refer to this as the AS model.

2.2.1 Solution

I will once again solve the model analytically using the method of undetermined coefficients. However, due to the latent variables z_t and g_t, this will be accomplished beginning with state space representation. First, note that the minimal solution of the model has the following general ABCD formulation (See Fernández-Villaverde et al. (2007) and Komunjer and Ng (2011)):

$$\begin{aligned}
\underbrace{\begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix}}_{X_t} &= \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{rz} & c_{rg} & c_{rr} \end{bmatrix}}_{A(C(\theta_s),\theta_s)} \underbrace{\begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d_{rz} & d_{rg} & d_{rr} \end{bmatrix}}_{B(D(\theta_s))} \underbrace{\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \end{bmatrix}}_{\varepsilon_t} \\
\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} &= \underbrace{\begin{bmatrix} c_{rz} & c_{rg} & c_{rr} \\ c_{yz} & c_{yg} & c_{yr} \\ c_{\pi z} & c_{\pi g} & c_{\pi r} \end{bmatrix}}_{C(\theta_s)} \begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} d_{rz} & d_{rg} & d_{rr} \\ d_{yz} & d_{yg} & d_{yr} \\ d_{\pi z} & d_{\pi g} & d_{\pi r} \end{bmatrix}}_{D(\theta_s)} \begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \end{bmatrix}
\end{aligned} \tag{27}$$

The scalars c_{ij} and d_{ij} are a-priori unknown functions of the structural parameters θ_s. The sizes of the vectors of observables Y_t , states X_t, and disturbances ε_t, are each 3. Note, the observables Y_t are the same as in the CS model.

First, it is useful to state a result which simplifies this representation. I present the proof pedantically in Appendix A.

Proposition 1. *The AS model has three analytical solutions, each $i = 1,\ 2,\ 3$ with ABCD representation*

$$\begin{aligned}
X_t &= A^{(i)}(C^{(i)}, \rho) \quad X_{t-1} + A^{(i)}\rho^{-1} \quad \varepsilon_t \\
Y_t &= \qquad C^{(i)} \qquad X_{t-1} + C^{(i)}\rho^{-1} \quad \varepsilon_t
\end{aligned} \tag{28}$$

for

$$C^{(i)}(\theta_s) = \begin{bmatrix} c_{rz}^{(i)} & 0 & c_{rr}^{(i)} \\ c_{yz}^{(i)} & \rho_g & c_{yr}^{(i)} \\ c_{\pi z}^{(i)} & 0 & c_{\pi r}^{(i)} \end{bmatrix} \qquad \rho = \begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ 0 & 0 & \rho_r \end{bmatrix} \tag{29}$$

Proposition 1 says that all elements of ABCD representation depend on $C^{(i)}$ and ρ. Furthermore, similar to the CS model, there are exactly three analytical solutions for each θ_s. However, this is not the only illuminating property of the model; as shown in Appendix Equation (A.4), one may define an invertible matrix F such that $Y_t = FX_t$. This expedites the following result, also proven in Appendix A.

Proposition 2. *Let Y_t have ABCD representation where ε_t is mean zero white noise with covariance Σ. If there exists an invertible matrix F such that $Y_t = FX_t$, then Y_t also has VAR(1) representation $Y_t = \Phi Y_{t-1} + U_t$ where $\Phi = CAC^{-1}$ and U_t is mean zero white noise with covariance $\Omega = D\Sigma D'$.*

Proposition 2 provides conditions for VAR(1) representation which are restrictive in the general sense of DSGE models, but useful in the current case (See Ravenna (2007) and Giacomini (2013)). Specifically, we have the following representation of the observables:

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \phi_{rr}^{(i)} & 0 & \phi_{r\pi}^{(i)} \\ \phi_{yr}^{(i)} & \rho_g & \phi_{y\pi}^{(i)} \\ \phi_{\pi r}^{(i)} & 0 & \phi_{\pi\pi}^{(i)} \end{bmatrix}}_{\Phi^{(i)}(\theta_s)} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt}^{(i)} \\ u_{yt}^{(i)} \\ u_{\pi t}^{(i)} \end{bmatrix}}_{U_t^{(i)}} \tag{30}$$

$U_t^{(i)} \sim NID(0, \Omega^{(i)}(\theta))$ is a vector of reduced form disturbances with unique elements

$$\text{vech}(\Omega^{(i)}) = \omega^{(i)} = D_3^+ (C^{(i)} \otimes C^{(i)})(\rho \otimes \rho)^{-1} D_3 \times \theta_\sigma \tag{31}$$

The nonzero elements of $\Phi^{(i)}$ are collected in the 7×1 vector $\phi^{(i)}$, with the relation

$$\text{vec}(\Phi^{(i)}) = (C^{(i)'-1} \otimes C)\text{vec}(A^{(i)}(C^{(i)}, \rho)) = R\phi^{(i)} \tag{32}$$

$$\underset{9\times 7}{R} = \left[\begin{array}{c:c:c} I_3 & 0 & 0 \\ \hdashline 0 & \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}' & 0 \\ \hdashline 0 & 0 & I_3 \end{array}\right] \tag{33}$$

Since $U_t^{(i)}$ is Gaussian, the reduced from parameters are

$$\underset{13\times 1}{\pi^{(i)}} = \left(\phi^{(i)\prime}, \omega^{(i)\prime}\right)' \tag{34}$$

Previously for the CS model, the mapping between $\phi^{(i)}$ and structural parameters θ_s, written $g^{\phi}(\phi^{(i)}, \theta_s) = 0$, was given in (9)-(11). This was easily obtained because $E_t Y_{t+1} = \phi r_t$ was taken as given. The AS model, however, is solved in Appendix A using the state space-implied condition $E_t Y_{t+1} = CY_t$, yielding solutions $g^{C}(C^{(i)}, \theta_s) = 0$. Since $\phi^{(i)}$ is also a known function of the elements of $C^{(i)}$, the relationship between $\pi^{(i)}$ and θ is in this case also closed-form. However, these indirect expressions prove too complicated to be useful towards providing intuition for identifiability, the ultimate objective.[10] This obstacle may be overcome, indirectly. Using the equality $E_t Y_{t+1} = \Phi^{(i)} Y_t$ now established by Proposition 2, we also have the following result:

Proposition 3. *The AS model has SVAR representation*

$$\Gamma_0^{(i)}(\phi^{(i)}, \theta_s) Y_t = \Gamma_1^{(i)}(\phi^{(i)}, \theta_s) Y_{t-1} + \varepsilon_t \tag{35}$$

where $\Gamma_0^{(i)} = \rho C^{(i)-1}$ *and* $\Gamma_0^{(i)-1}\Gamma_1^{(i)} = \Phi^{(i)}$.

Unlike $C^{(i)}$, both $\Gamma_0^{(i)}$ and $\Gamma_1^{(i)}$ – given in the proof to Proposition 3 in Appendix A – are relatively simple functions of the reduced form parameters and structural parameters. Thus, they may be used to produce a sensible mapping between θ and $\phi^{(i)}$. This is defined in parts by

$$g^{\phi}(\phi^{(i)}, \theta_s) = \text{vec}(\Gamma_1^{(i)}) - (I_3 \otimes \Gamma_0^{(i)}) R\phi^{(i)} = 0_{7\times 1} \tag{36}$$

$$g^{\omega}(\pi^{(i)}, \theta) = \theta_\sigma - D_3^{+}(\Gamma_0^{(i)} \otimes \Gamma_0^{(i)}) D_3 \times \omega^{(i)} = 0_{6\times 1} \tag{37}$$

In other words, $\left[g^{\phi\prime} \quad g^{\omega\prime}\right]' = 0$ makes up the 13×1 moment condition (1) to be utilized henceforth.

2.2.2 Determinacy and indeterminacy

Finally, as in the case of the CS model, the fact that there are three solutions for each θ does not imply indeterminacy for all points. The eigenvalues $\{\lambda\}$ of $\Phi^{(i)}$ are defined

[10] Each element of $C^{(i)}$ takes many pages of text to express. See `uap/run/as/solve/solns`.

as the solutions to the characteristic equation

$$(\lambda - \rho_g)((\lambda - \phi_{rr}^{(i)})(\lambda - \phi_{\pi\pi}^{(i)}) - \phi_{r\pi}^{(i)}\phi_{\pi r}^{(i)}) = 0 \tag{38}$$

The solutions are thus ρ_g and the roots of $\lambda^2 - (\phi_{rr} + \phi_{\pi\pi})\lambda + (\phi_{rr}\phi_{\pi\pi} - \phi_{r\pi}\phi_{\pi r}) = 0$. In Figure 2, I compare solutions at two points, corresponding to active monetary policy θ_0, and passive monetary policy θ_1. Again, for this model I find that the solution is indeterminate under passive, but not active monetary policy. Finally, Table 2 again also serves to prove the analytical solutions are correct. At θ_0, Solution 1 is exactly that provided by Sims (2002)'s `gensys` routine. At θ_1, Sims's routine returns the error message, "`Indeterminacy.  1 loose endog errors.`"

3 Identifiability

In the previous section I have shown that the CS and AS models each have three analytical solutions, which need not be either stable or real-valued, and restricted VAR(1) reduced form representation. I have also defined a closed-form mapping from structural to reduced form parameters g, correponding to Equation (1), in each case.

In this section, I consider identifiability in each model on the basis of each g. In the case of the CS model, there were 9 reduced form parameters and 11 structural parameters. For AS, there were 13 reduced form and 14 structural. Thus, the necessary order condition is not met, and neither model is identifiable without further restriction. In such situations, the most common contemporary approach is to fix some parameters equal to constants, in hope of identifying their complement. This practice is more formally known as *conditional identification*. Since conditional identification is known to potentially produce biases (See Canova and Sala (2009)), I will consider the case study of an analyst who rationally fixes as few parameters as possible: 2 for CS and 1 for AS. The criterion she will use to choose these parameters will be currently available results on local identifiability. Do such restrictions also ensure global identifiability?

Recall from the results of Rothenberg (1971) that a parameter point θ is said to be *locally* identifiable in an ϵ-neighborhood of θ if in that neighborhood there exists no $\theta^* \neq \theta$ such that θ and θ^* are observationally equivalent. More generally, θ is said to be *globally* identifiable in the parameter space Θ if there is no other $\theta^* \in \Theta$

Determinacy: Active monetary policy.

θ_0 :

τ	β	κ	ψ_π	ψ_y	ρ_z	ρ_g
2	0.9975	0.33	1.5	0.125	0.9	0.95
ρ_r	$100\sigma_z$	$100\sigma_g$	$100\sigma_r$	ρ^σ_{gz}	ρ^σ_{rz}	ρ^σ_{rg}
0.75	0.3	0.6	0.2	0.5	-0.5	-0.5

Solution i	$\Phi^{(i)}$	1e5 $\times\ \Omega^{(i)}$	λ	Stable?
1	$\begin{bmatrix} 0.69 & 0 & 0.50 \\ 0.14 & 0.95 & -0.34 \\ -0.24 & 0 & 1.46 \end{bmatrix}$	$\begin{bmatrix} 1.91 & \cdot & \cdot \\ -4.41 & 13.7 & \cdot \\ 3.13 & -7.88 & 5.5 \end{bmatrix}$	$\begin{bmatrix} 1.25 \\ 0.95 \\ 0.9 \end{bmatrix}$	no
2	$\begin{bmatrix} 0.69 & 0 & 0.48 \\ 0.08 & 0.95 & -0.19 \\ -0.21 & 0 & 1.38 \end{bmatrix}$	$\begin{bmatrix} 1.55 & \cdot & \cdot \\ -2.93 & 8.66 & \cdot \\ 2.25 & -4.71 & 3.55 \end{bmatrix}$	$\begin{bmatrix} 1.17 \\ 0.95 \\ 0.9 \end{bmatrix}$	no
3	$\begin{bmatrix} 0.79 & 0 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0 & 0.62 \end{bmatrix}$	$\begin{bmatrix} 0.27 & \cdot & \cdot \\ 0.54 & 11.05 & \cdot \\ 0.44 & 4.94 & 2.89 \end{bmatrix}$	$\begin{bmatrix} 0.95 \\ 0.9 \\ 0.51 \end{bmatrix}$	yes

Indeterminacy: Passive monetary policy.

θ_1 : Same as θ_0, but with $\psi_\pi = 0.1$.

Solution i	$\Phi^{(i)}$	1e5 $\times\ \Omega^{(i)}$	λ	Stable?
1	$\begin{bmatrix} 1.01 & 0 & -0.02 \\ 3.31 & 0.95 & -0.70 \\ -2.4 & 0 & 1.40 \end{bmatrix}$	$\begin{bmatrix} 1.4 & \cdot & \cdot \\ 42.6 & 1369 & \cdot \\ -0.028 & -902 & 596 \end{bmatrix}$	$\begin{bmatrix} 1.51 \\ 0.95 \\ 0.9 \end{bmatrix}$	no
2	$\begin{bmatrix} 0.76 & 0 & 0.03 \\ -0.18 & 0.95 & 0.03 \\ -0.24 & 0 & 0.95 \end{bmatrix}$	$\begin{bmatrix} 0.49 & \cdot & \cdot \\ -0.14 & 2.60 & \cdot \\ 1.15 & -0.63 & 3.28 \end{bmatrix}$	$\begin{bmatrix} 0.95 \\ 0.9 \\ 0.81 \end{bmatrix}$	yes
3	$\begin{bmatrix} 0.74 & 0 & 0.03 \\ 0.75 & 0.95 & -0.17 \\ 0.58 & 0 & 0.78 \end{bmatrix}$	$\begin{bmatrix} 0.32 & \cdot & \cdot \\ -1.36 & 8.85 & \cdot \\ -0.33 & 2.94 & 3.8 \end{bmatrix}$	$\begin{bmatrix} 0.95 \\ 0.9 \\ 0.62 \end{bmatrix}$	yes

Table 2: Determinacy and indeterminacy in AS. ρ^σ_{ij} denotes the correlation $\sigma_{ij}/(\sigma_i\sigma_j)$.

such that $\theta^* \neq \theta$ but θ and θ^* are observationally equivalent. Thus, it is evident that restrictions guaranteeing local identifiability need not ensure global, since nuisance θ^* may exist outside of an ϵ-neighborhood of θ. Yet, global identifiability in particular can be difficult to prove, particularly for nonlinear models, like DSGEs.[11] The reason why is that the lynchpin of both definitions of global and local identifiability is *observational equivalence*. Two points are said to be observationally equivalent if and only if the likelihood function evaluated at either is the same for any data set. Consider the brute-force search necessary, for instance, to find two parameter values yielding the same value of the likelihood for even one data set. It is generally not possible to find all observationally equivalent points in an unbounded parameter space, like the real numbers.

Yet, because of the reduced form representation proven in the previous section, we also have the immediate Corollary to Rothenberg's results:[12]

Corollary 1. *Let $\pi^{(i)}(\theta)$ be a reduced form parameter. Two structural parameters θ and θ^* are observationally equivalent if and only if $\pi^{(i)}(\theta) = \pi^{(j)}(\theta^*)$ for i possibly not, though not necessarily exclusive of, j.*

The analytical tractability of g developed in the previous section makes both local and global identifiability easy to assess for any parameter space under Corollary 1. Furthermore, these results are robust to unbounded ranges of parameter values, such as the real numbers, rather than specific points or regions.

3.1 CS identifiability

The CS model has 9 reduced form parameters in $\pi^{(i)}$ versus 11 structural parameters in θ. The necessary order condition for identifiability is therefore not satisfied. Thus, let us consider the case of an analyst who, armed only with results on local identifiability, chooses two parameters to fix to constants.

Table 3 depicts the identification pathologies in the CS model graphically.[13] Accordingly to the respective dimensions of $\pi^{(i)}$ and θ, the table has 9 rows and 11 columns.

[11] Nonlinear in the sense of the mapping from structural parameters to likelihood function.

[12] See Rothenberg (1971) p. 578 Definitions 1-3 and pp. 584-5 assumptions VII and VIII.

[13] The mapping is computed in closed form by `makeCS.m` in the attached documentation. I have adopted the idea of this diagram from the identification tables studied by Hamilton and Wu (2012) with respect to affine term structure models. A distinction is the use of the digit 2 in use here, where global identifiability is specifically the issue of concern.

		θ_s					θ_σ					
		τ	β	κ	ψ_π	ρ_r	σ_r^2	σ_{yr}	$\sigma_{\pi r}$	σ_y^2	$\sigma_{\pi y}$	σ_π^2
g^ϕ	g_1^ϕ	✓										
	g_2^ϕ		✓	✓								
	g_3^ϕ				✓	✓						
g^ω	g_1^ω					2	✓					
	g_2^ω					2		✓				
	g_3^ω			✓		2			✓			
	g_4^ω					2				✓		
	g_5^ω			✓		2					✓	
	g_6^ω			2		2						✓

Table 3: Identifiability in CS model on the basis of g.

A single checkmark ✓ in a row represents the fact that a parameter is identifiable from that degree of freedom. For instance, τ is uniquely identifiable from the first row of g^ϕ, labeled g_1^ϕ. This is directly verifiable from (9). When two checkmarks appear in a given row, then those parameters are not separately identifiable on the basis of that degree of freedom. For instance, β and κ are linearly dependent on the basis of the second row of g^ϕ, labeled g_2^ϕ. However, either is uniquely identifiable conditionally on the other. These points are directly verifiable from (10). When a "2" appears in an entry, however, there are two equally probable values of that parameter on the basis of the degree of freedom. For instance, ρ_r and σ_r^2 are linearly dependent on the basis of the first row of g^ω, labeled g_1^ω. However, whereas σ_r^2 is uniquely identifable conditionally on ρ_r, there are two separately unidentifable values of ρ_r which are equally likely on the basis of g_1^ω, conditionally on σ_r^2.

The cause of the digits "2" for both κ and ρ_r is relatively transparent in the current setting; since $\Omega^{(i)} = \delta^{(i)}\Sigma\delta^{(i)\prime}$, the appearance of each κ and ρ_r in $\delta^{(i)}$ leads the reduced form parameters $\omega^{(i)} = \text{vech}(\Omega^{(i)})$ to be quadratic functions of κ and ρ_r. However, more generally the mapping from structural parameters to likelihood in DSGE models is a black box, and currently available results on local identifiability provide no criterion for choosing conditional identification scheme. What would be the outcome for an analyst blindly choosing conditional identification schemes in this case? In Table 4, I consider all possible 35 choices of two parameters which ensure at least local identifiability of the complement set. The results are striking. In 77% of all cases, the analyst using only results on local identifiability will not also obtain global identifiability.

Global	Local only
β or κ, ψ_π or ρ_r	β or κ, Any θ_σ
$\sigma_{\pi r}$ or $\sigma_{\pi y}$, ψ_π or ρ_r	$\sigma_{\pi r}$, $\sigma_{\pi y}$, or σ^2_π, Any other θ_σ
8 =23% Cases	27 = 77% Cases

Table 4: Identifiability consequences of conditional identification schemes in CS.

The consequences of the above findings of Table 4 are best illustrated with a numerical example. Consider the analyst who fixes the parameters β and σ_r to constants, and recall the determinate value θ_0 listed in Table 1. According to Table 3, θ_0 has exactly one observationally point. This observationally equivalent point, labeled θ_0^*, is given in Table 5.[14] First, note that only the parameter values $(\psi_\pi, \rho_r, \rho^\sigma_{yr}, \rho^\sigma_{\pi r})$ differ parameter-to-parameter. Thus, a subset of θ is globally identifiable. Second, note that it is Solution 3 in Table 5 which is the same as Solution 1 in Table 1. Thus, all possible solutions must be considered. Third, note that the point θ_0^*, unlike θ_0, is indeterminate. Thus, one must account for indeterminacy or risk overlooking observationally equivalent points. Finally, while θ_0 is relatively consistent with usual priors from macroeconomic theory, θ_0^* is not. For example, $\psi_\pi < 0$. I will return to this observation in the next section.

The fact that the reduced form parameters are by definition the same for observationally equivalent points, and that a subset of θ is globally identifiable in the above example, may lead one to guess that the distinction between θ_0 and θ_0^* is not substantive from a theoretic or policy perspective; is it not true that observationally equivalent points imply similar macroeconomic dynamics? This misconception is best addressed by considering the monetary policy impulse-responses at either point, depicted in Figure 1. Whereas contractionary monetary policy leads to decreased output and inflation at θ_0, it is expansionary at θ_0^*. The algebraic explanation is relatively simple: While $\pi^{(i)}(\theta_0) = \pi^{(i)}(\theta_0^*)$ by definition, $\delta^{(i)}(\theta_0) \neq \delta^{(i)}(\theta_0^*)$ are not reduced form parameters, but are necessary to compute impulse-responses. However, the implications for how to proceed are at the moment at least unclear.

For this case, I have realistically assumed that the analyst has no means of ranking conditional identification schemes a-priori. In response to this, one might still argue, a choice of fixing both β and ψ_π is typical, because typical priors regarding these param-

[14]These observatioinally equivalent values are computed by `piToThetaOECS.m`.

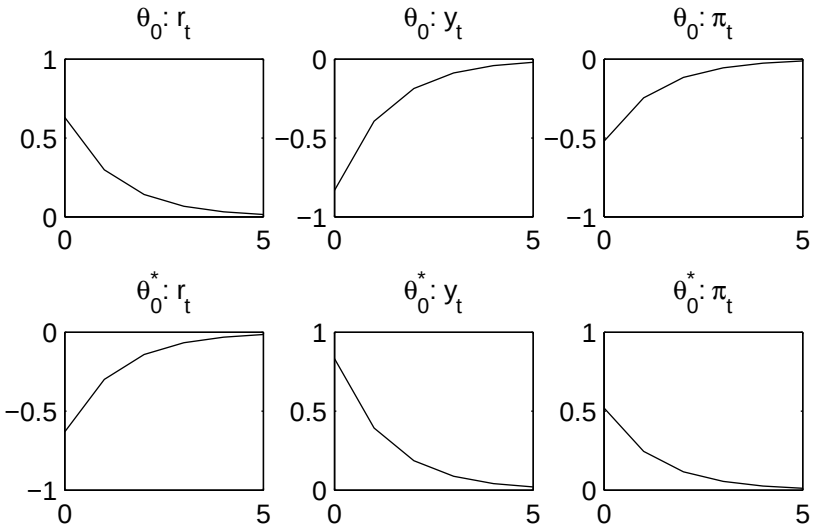
θ_0^* :

τ	β	κ	$\boldsymbol{\psi_\pi}$	$\boldsymbol{\rho_r}$	$100\sigma_r$	$100\sigma_y$	$100\sigma_\pi$	$\boldsymbol{\rho^\sigma_{yr}}$	$\boldsymbol{\rho^\sigma_{\pi r}}$	$\rho^\sigma_{\pi y}$
2	0.9975	0.33	**-6.63**	**-0.75**	0.2	0.2	0.2	**-0.5**	**-0.5**	-0.5

Solution i	$\begin{bmatrix}\phi^{(i)} & 0\end{bmatrix}$	1e5 $\times$ $\Omega^{(i)}$	Stable?
1	$\begin{bmatrix} 2.64 & 0 & 0 \\ 0.96 & 0 & 0 \\ -0.19 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 4.97 & \cdot & \cdot \\ 2.51 & 1.57 & \cdot \\ 0.57 & 0.09 & 0.2 \end{bmatrix}$	no
2	$\begin{bmatrix} -0.6 & 0 & 0 \\ 0.18 & 0 & 0 \\ 0.04 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.26 & \cdot & \cdot \\ -0.24 & 0.52 & \cdot \\ -0.23 & 0.01 & 0.34 \end{bmatrix}$	yes
3	$\begin{bmatrix} 0.47 & 0 & 0 \\ -0.62 & 0 & 0 \\ -0.39 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.16 & \cdot & \cdot \\ -0.08 & 0.34 & \cdot \\ 0.03 & -0.22 & 0.14 \end{bmatrix}$	yes

Table 5: Observational equivalence in CS: θ_0^* is indeterminate observationally equivalent point to determinate θ_0 given in Table 1. **Bold** parameters are distinct from θ_0.

Figure 1: CS monetary policy impulse-responses: θ_0 vs. θ_0^*



eters are tight, and therefore global identifiability is not a concern. I now disprove this argument by showing similar conditional identification restrictions for the AS model never imply global identifiability.

3.2 AS identifiability

The AS model has 13 reduced form parameters and 14 structural parameters. As in the CS model, all structural parameters are therefore not identifiable. Table 6 depicts the identification pathologies in AS graphically.[15]

Table 6 is similar to Table 3 with one important exception; in this case, it is occasionally necessary to use two implicit functions from g^ϕ to solve for two elements of θ_s jointly. For instance, a checkmark in row g_4^ϕ column ρ_g still indicates ρ_g is identified uniquely by the fourth row of g^ϕ, labeled g_4^ϕ. However, a checkmark in row (g_2^ϕ, g_6^ϕ) column (β, κ) indicates that β and κ are jointly uniquely identifiable by g_2^ϕ and g_6^ϕ; in other words, each g_2^ϕ and g_6^ϕ is a function of both β and κ, and solving this system of two equations and two unknowns recovers β and κ uniquely. However, the 2 in row (g_1^ϕ, g_5^ϕ) in column (τ, ρ_z) signals another identifiability problem. These two equations may be used to solve for both τ and ρ_z, but the solution yields two values of each. In other words, τ and ρ_z are locally, but not globally identifiable on the basis of g_1^ϕ and g_5^ϕ. Finally, a checkmark or 2 indexed with dimensions represents a matrix, for example $\checkmark_{2\times 1} = \begin{bmatrix}\checkmark & \checkmark\end{bmatrix}'$. For example, $\checkmark_{2\times 1}$ in row (g_2^ϕ, g_6^ϕ) indicates that ρ_g is uniquely identifiable using either g_2^ϕ or g_6^ϕ. $\checkmark_{2\times 3}$ in row (g_3^ϕ, g_7^ϕ) column $(\psi_y, \psi_\pi, \rho_r)$ indicates that any two elements of the Taylor rule are identifiable on the basis of those equations, but all three are together underidentified.

The minimum amount of parameters that must be fixed to ensure local identifiability in the AS model is one. Since all three parameters of the Taylor rule are not identifiable, that means that one must fix one of $(\psi_y, \psi_\pi, \rho_r)$ or $(\sigma_{rz}, \sigma_{rg}, \sigma_r^2)$ to ensure local identifiability. However, none of these restrictions address the two solutions for (τ, ρ_z). Thus, in none of these cases is global identifiability implied. Therefore, the analyst who fixes only one parameter to ensure local identifiability will in no case yield global identifiability. In Table 7 I list the single observational equivalent point to θ_0 in Figure 2. As was the case for the CS model, θ_0^* is indeterminate, and possibly

[15]The mapping described by Table 6 is computed in closed form by `makeAS.m`.

		θ_s				θ_σ					
		ρ_g	(β, κ)	(τ, ρ_z)	$(\psi_y, \psi_\pi, \rho_r)$	σ_z^2	σ_{gz}	σ_{rz}	σ_g^2	σ_{rg}	σ_r^2
g^ϕ	g_4^ϕ	✓									
	(g_2^ϕ, g_6^ϕ)	$\checkmark_{2\times1}$	✓								
	(g_1^ϕ, g_5^ϕ)	$\checkmark_{2\times1}$	✓	2							
	(g_3^ϕ, g_7^ϕ)		✓		$\checkmark_{2\times3}$						
g^ω	g_1^ω	2	$2_{1\times2}$	$2_{1\times2}$		✓					
	g_2^ω	✓	$2_{1\times2}$	$\checkmark_{1\times2}$			✓				
	g_3^ω	✓	$2_{1\times2}$	$\checkmark_{1\times2}$	$\checkmark_{1\times3}$			✓			
	g_4^ω		$2_{1\times2}$						✓		
	g_5^ω		$2_{1\times2}$		$\checkmark_{1\times3}$					✓	
	g_6^ω		$2_{1\times2}$		$2_{1\times3}$						✓

Table 6: Identifiability in AS model on the basis of g.

less a-priori likely than θ_0 on the basis of usual priors from macroeconomic theory. For example, CRRA $\tau < 0$. In Figure 2 I depict monetary policy impulse-responses. Contractionary policy at θ_0 is once again paradoxically expansionary at θ_0^*.

Finally, thus far I have only considered data which is mean-zero. A necessary robustness check is that the problems pertaining to global identifiability I have outlined are not ameliorated simply by including means. In Appendix B.2 I reconsider the AS model using means of the data, and document exactly similar results in Table B.2.

4 Maximum likelihood estimation

In the previous section, I have shown that the CS and AS models are not generally globally identifiable on the basis of popularly utilized restrictions for local identifiability. This is problematic, because even the direction of monetary policy action may not be pinned down on the basis of the data. Consequent of this observational equivalence is also that the maximum likelihood estimator is not consistent, and there exists no theory from which to interpret any statistic from a Bayesian posterior as asymptotically unbiased.

In this section, I consider the implications for the maximum likelihood estimator when parameters are not globally identifiable. This proceeds in two parts. First, I derive a "plug-in" maximum likelihood estimator, which is based on estimating $\widehat{\pi}$ by least squares, and then infering the estimates $\widehat{\theta}$ indirectly using the analytical mapping g. With no need for numerical search, this estimator is extremely computationally

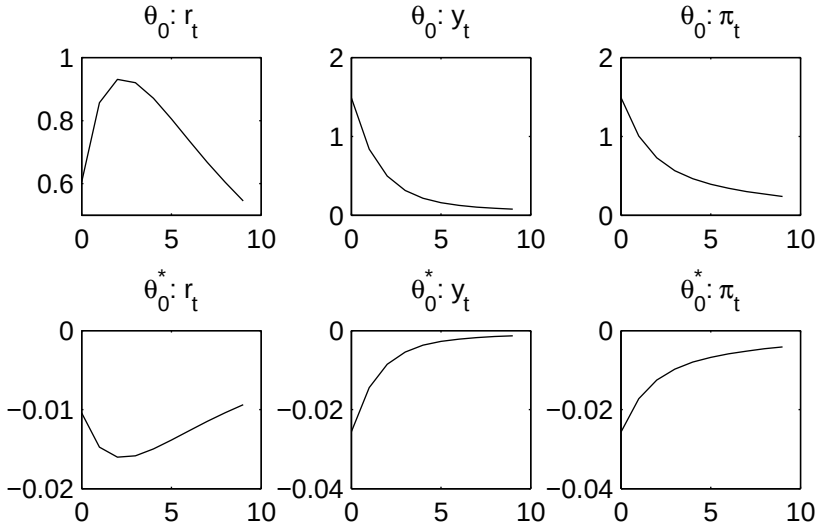
θ_0^* :

$\boldsymbol{\tau}$	β	κ	ψ_π	ψ_y	$\boldsymbol{\rho_z}$	ρ_g
$\mathbf{-45.45}$	0.9975	0.33	1.5	0.125	$\mathbf{0.51}$	0.95
ρ_r	$\mathbf{100\sigma_z}$	$100\sigma_g$	$100\sigma_r$	$\boldsymbol{\rho^\sigma_{gz}}$	$\boldsymbol{\rho^\sigma_{rz}}$	ρ^σ_{rg}
0.75	$\mathbf{23.94}$	0.6	0.2	$\mathbf{-0.57}$	$\mathbf{0.78}$	-0.5

Solution i	$\Phi^{(i)}$	$1\text{e}5 \times \Omega^{(i)}$	λ	Stable?
1	$\begin{bmatrix} 0.83 & 0 & 0.29 \\ 0.28 & 0.95 & -0.38 \\ 0.24 & 0 & 0.73 \end{bmatrix}$	$\begin{bmatrix} 4.71 & \cdot & \cdot \\ 9.3 & 9.36 & \cdot \\ 1.14 & 5.82 & 5.26 \end{bmatrix}$	$\begin{bmatrix} 1.05 \\ 0.95 \\ 0.41 \end{bmatrix}$	no
2	$\begin{bmatrix} 0.79 & 0 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0 & 0.62 \end{bmatrix}$	$\begin{bmatrix} 0.29 & \cdot & \cdot \\ 0.54 & 11.1 & \cdot \\ 0.44 & 4.94 & 2.89 \end{bmatrix}$	$\begin{bmatrix} 0.95 \\ 0.9 \\ 0.51 \end{bmatrix}$	yes
3	$\begin{bmatrix} 0.76 & 0 & 0.23 \\ 0.08 & 0.95 & -0.57 \\ 0.04 & 0 & 0.55 \end{bmatrix}$	$\begin{bmatrix} 0.18 & \cdot & \cdot \\ 0.12 & 10.7 & \cdot \\ 0.11 & 3.79 & 1.74 \end{bmatrix}$	$\begin{bmatrix} 0.95 \\ 0.8 \\ 0.51 \end{bmatrix}$	yes

Table 7: Observational equivalence in AS: θ_0^* is indeterminate observationally equivalent point to determinate θ_0 given in Table 2. **Bold** parameters are distinct from θ_0.

Figure 2: AS monetary policy impulse-responses: θ_0 vs. θ_0^*



efficient, taking less than a second to obtain the MLE, even when multiplicatively-valued. Second, I draw on the computational efficiency of the plug-in MLE to compute the small sample distribution with, and without, global identifiability.

4.1 Plug-in MLE

Hamilton and Wu (2012) consider identification and estimation in affine term structure models. The models they consider, similar to the current paper, have Gaussian reduced form representation. They show that for full information maximum likelihood estimators of the reduced form parmeters $\widehat{\pi}$ of a given model, one may obtain an asymptotically efficient minimum chi-squared estimator (MCSE) of the structural parameters $\widehat{\theta}_{MCSE}$ by minimimizing the criterion

$$(\widehat{\pi} - \pi(\theta))'\widehat{\mathcal{I}}(\widehat{\pi} - \pi(\theta)) \tag{39}$$

where $\widehat{\mathcal{I}}$ is a consistent estimator of the information matrix with respect to π. This estimator is useful, since for the models they consider, $\widehat{\pi}$ is often available from OLS, and $\widehat{\mathcal{I}}$ is closed-form. Furthermore, in the just-identified case, the minimum possible value of the MCSE criterion is zero. As the authors explain, one may then simply minimize

$$(\widehat{\pi} - \pi(\theta))'(\widehat{\pi} - \pi(\theta)) \tag{40}$$

In this latter case, the MCSE is also numerically equivalent to MLE. This feature is useful because one will know with certainty when the maximum likelihood estimator is obtained by simply checking that the criterion is zero.

Both the CS and AS models introduced in the previous sections may now be represented by

$$Y_t = (Y_{t-1}' \otimes I_3)R\phi^{(i)} + U_t^{(i)} \tag{41}$$

where R is a zero-one selection matrix, $U_t^{(i)}$ is iid Gaussian with unique covariances $\omega^{(i)}$, i is the index of the solution, and the reduced form parameters are $\pi^{(i)} = \left(\phi^{(i)\prime}, \omega^{(i)\prime}\right)'$.[16] When, as motivated in the previous section, the analyst fixes 2 parameters for CS, or one 1 for AS, she is also considering a just-identified model. Thus,

[16]The representation (41) comes from the fact that $\text{vec}(\Phi^{(i)}Y_{t-1}) = (Y_{t-1}' \otimes I_3)\text{vec}(\Phi^{(i)})$ and $\text{vec}(\Phi^{(i)}) = R\phi^{(i)}$. R has already been described for the AS model: See (33). For the CS model it is $R = \begin{bmatrix} I_3 & 0_{3\times 6} \end{bmatrix}'$.

extrapolating on Hamilton and Wu's approach, $\widehat{\pi}$ may simply be obtained by feasible generalized least squares (FGLS), and the criterion (40) may be used to obtain $\widehat{\theta}_{MLE}$.[17]

Yet, in the current situation, there is an additional advantage to using (40); because of the availability of a closed form moment condition for the reduced form paramaters g, one need not utilize any numerical minimization to obtain the MLE. Rather, one may simply estimate $\widehat{\pi}_{FGLS}$ and invert g with respect to θ to obtain $\widehat{\theta}_{MLE}$. In other words, if θ is globally identifiable, then the maximum likelihood estimator is given by

$$g(\widehat{\pi}_{FGLS}, \theta)|_{\theta=\widehat{\theta}_{MLE}} = 0 \;\rightarrow\; \widehat{\theta}_{MLE} = g^{-1}(0; \widehat{\pi}_{FGLS}) \tag{42}$$

This estimator is also a special case of indirect least squares. Possibly more interesting, however, is what happens when global identifiability is not ensured. For example, if every parameter value has one observationally equivalent value, then g^{-1} will be multiplicatively-valued, yielding

$$g^{-1}(0; \widehat{\pi}_{FGLS}) = \begin{cases} \widehat{\theta}_{MLE} \\ \widehat{\theta}^{*}_{MLE} \end{cases} \tag{43}$$

for $\widehat{\theta}_{MLE} \neq \widehat{\theta}^{*}_{MLE}$. Thus, g enables one to compute all modes of the likelihood for a given data sample without the need for multiple numerical searches. Before putting this advantageous quality to work in the case where parameters are not globally identifiable, I demonstrate that this estimator is in fact the MLE as claimed in an example with global identifiability.

4.1.1 MLE with global identifiability

The FGLS estimator $\widehat{\pi}_{FGLS}$ and bootstrap methodology for small sample confidence interval construction for $\widehat{\theta}_{MLE}$ are detailed in Appendix C. Consider the local identifying restriction that β and ψ_π are fixed to their true values, called "Scheme A" in Table 8, implying global identifiability. For data-generating θ_0 as given in Table 1, I generate a time series of length $T = 250$ and utilize the plug-in MLE. $T = 250$ represents approximately the dimensions of post-war quarterly data regularly utilized for DSGE estimation. Table 8 gives the result of this estimation, with estimate labeled $\widehat{\theta}_A$, and

[17] GLS is equivalent to MLE under normality. See Hamilton (1994) p. 222. FGLS is in this case necessary because the covariances ω are unknown.

Scheme A: Fix β and ψ_π to their true values in θ_0.

	τ	κ	ρ_r	$100\sigma_r$	$100\sigma_y$	$100\sigma_\pi$	ρ^σ_{yr}	$\rho^\sigma_{\pi r}$	$\rho^\sigma_{\pi y}$
$\widehat{\theta}_A$	1.71	0.43	0.67	0.21	0.21	0.22	0.57	0.45	−0.47
upper	2.5	0.66	0.86	0.23	0.25	0.27	0.8	0.56	−0.27
lower	0.66	0.03	0.47	0.18	0.16	0.14	0.37	0.35	−0.7

Likelihood contours: $\ell = 5847.3$ at mode.

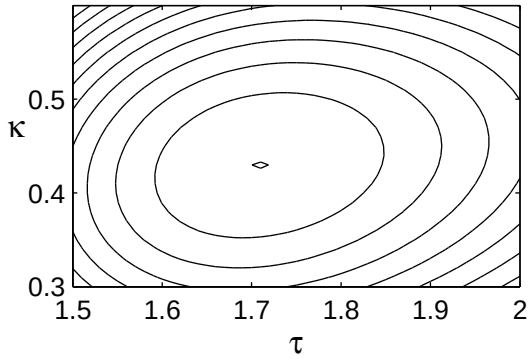


Table 8: Plug-in MLE and 95% confidence interval: Scheme A. Data-generating θ_0, $T = 250$.

upper and lower bounds for 95% bootstrap confidence intervals. FGLS estimators of the reduced form parameters are also given, in Appendix Table C.2.

The primary advantages of the plug-in MLE are its joint numerical certainty and computational efficiency; each estimate takes under a second to compute, allowing said bootstrap to compute in a matter of minutes. To verify that these estimates are truly equal to the MLE, I also calculate the Hessian of the model's log-likelihood, which is written

$$\ell(Y;\theta) = -\frac{3T}{2}\ln(2\pi) + \frac{T}{2}\ln|\Omega^{-1}| - \frac{1}{2}\sum_{t=2}^{T}(Y_t - \phi r_{t-1})'\Omega^{-1}(Y_t - \phi r_{t-1}) \tag{44}$$

The likelihood obtains a value of $\ell = 5847.3$ at $\widehat{\theta}_A$ and the Hessian is negative definite, indicating that plug-in MLE is truly the MLE as claimed. The largest of all negative eigenvalues of the Hessian is -54,126, indicating that there is no flatness in the likelihood, as sometimes associated with weak identifiability (See Canova and Sala (2009) and Qu (2014)). Table 8 also verifies that the plug-in MLE obtains the likelihood maximum graphically for one cross-section. The contours of the likelihood peak at exactly the values yielded by $\widehat{\theta}_A$.

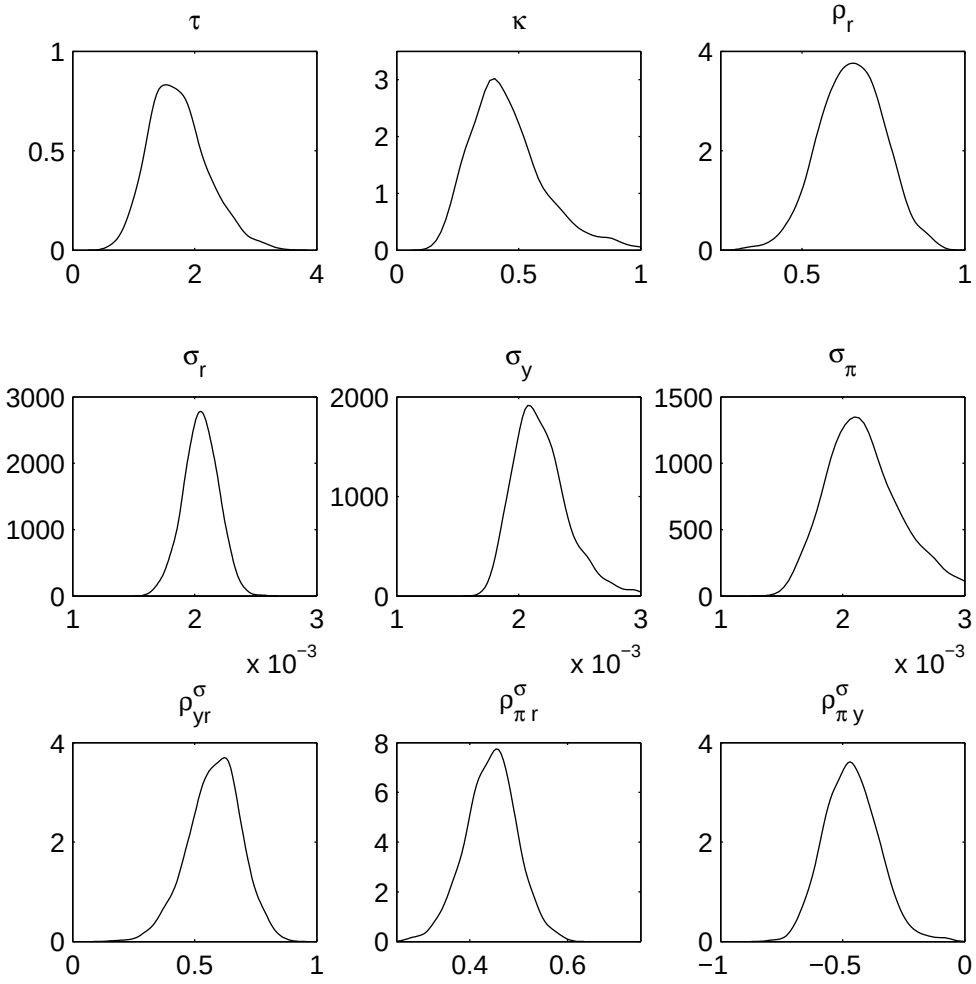
Finally, in Figure 3 I depict the kernel smoothing function estimates of the MLE's small sample distribution. Evidently, the distributions do not differ markedly from normal, and taken together with the previously demonstrated negative definiteness of the Hessian, provide no clear evidence of weak identifiability. However, in the next section, I consider how this and the previous findings differ when the same estimation is applied to a local identifiability scheme which does yield global identifiability.

4.2 MLE without global identifiability

Previously I have shown that the plug-in MLE is indeed the MLE, and that the MLE has standard small sample distributional properties, with no clear evidence of weak identifiability. In Table 9, I consider estimation using the same previously constructed synthetic data set, but under the assumption that instead β and σ_r are set to their true values in θ_0, called Scheme B. As explained previously, this conditional identification scheme implies local identifiability of the complement subset of θ, but not global identifiability. The material implication of this development is that the MLE is now set-valued, $\{\widehat{\theta}_B, \widehat{\theta}^*_B\}$. In Table 9, dashes indicate that the given estimates for $\widehat{\theta}^*_B$ are the same as $\widehat{\theta}_B$. However, for the subset of parameters $(\psi_\pi, \rho_r, \rho^\sigma_{yr}, \rho^\sigma_{\pi r})$ the estimates differ. As was the case for the data-generating value θ_0 and its observationally equivalent point θ^*_0, note that θ^*_B is less consistent with priors from macroeconomic theory than θ_B. To verify these computations are correct, I evaluate the likelihood at both $\widehat{\theta}_B$ and $\widehat{\theta}^*_B$ and obtain the same maximum value as for Scheme A, $\ell = 5847.3$. Furthermore, the Hessian is negative definite at both modes, and the largest eigenvalue of either Hessian is -55,665, in contrast with evidence of weak identifiability. The likelihood contours depict multi-modalness in one cross-section; both modes obtain exactly the same likelihood maximum as did the single mode under Scheme A.

The fact that θ_0 has an observationally equivalent point θ^*_0 under Scheme B means that the MLE is not consistent, and this is the underlying cause of the multi-valued estimates in Table 9. Consistency, in turn, is a formal requirement for asymptotic normality of the MLE. Therefore, it is no longer the case that the small sample distribution of the MLE will have the same usual bell-shape displayed in Figure 3 for Scheme A. In Figure 4, I display the kernel density estimates of $\{\widehat{\theta}_B, \widehat{\theta}^*_B\}$. Intuitively, the distribution is multimodal in the dimension of the four parameters which differed

Figure 3: Distribution of $\widehat{\theta}_A$.



Scheme B: Fix $\boldsymbol{\beta}$ and $\boldsymbol{\sigma_r}$ to their true values in $\boldsymbol{\theta_0}$.

	τ	κ	ψ_π	ρ_r	$100\sigma_y$	$100\sigma_\pi$	ρ^σ_{yr}	$\rho^\sigma_{\pi r}$	$\rho^\sigma_{\pi y}$
$\widehat{\theta}_B$	1.71	0.43	1.39	0.65	0.21	0.22	0.57	0.45	−0.47
$\widehat{\theta}^*_B$	–	–	−6.19	−0.65	–	–	−0.57	−0.45	–
upper	2.5	0.66	3.28	0.8	0.25	0.27	0.75	0.53	−0.27
lower	0.66	0.03	−6.8	−0.8	0.16	0.14	−0.75	−0.53	−0.7

Likelihood contours. $\ell = 5847.3$ at both modes.

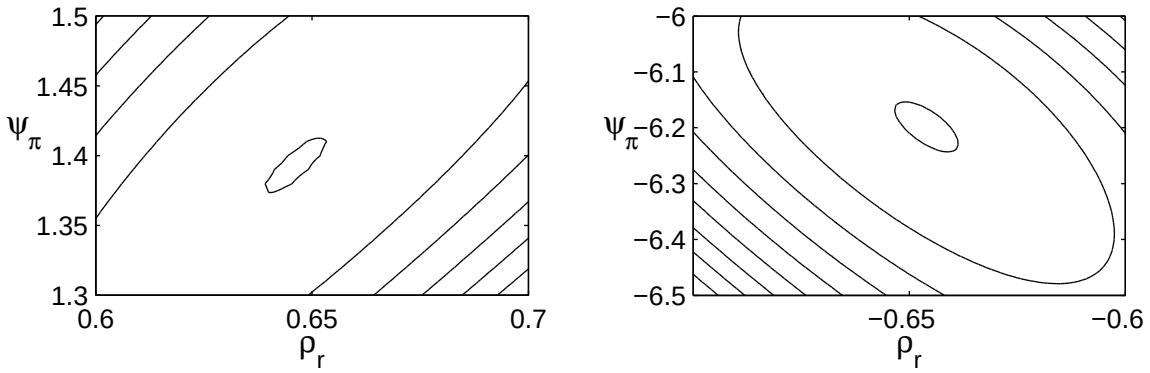


Table 9: Plug-in MLE and 95% confidence interval: Scheme B. Data-generating θ_0, $T = 250$.

from $\widehat{\theta}_B$ to $\widehat{\theta}^*_B$, $(\psi_\pi, \rho_r, \rho^\sigma_{yr}, \rho^\sigma_{\pi r})$, piling up at each respective mode. From our perspective, it is clear that this non-normality is caused by observational equivalence, but to the contemporary analyst, this might otherwise be misinterpreted as evidence of weak identifiability. As is clear from Scheme A, there is no evidence of weak identifiability in this model.

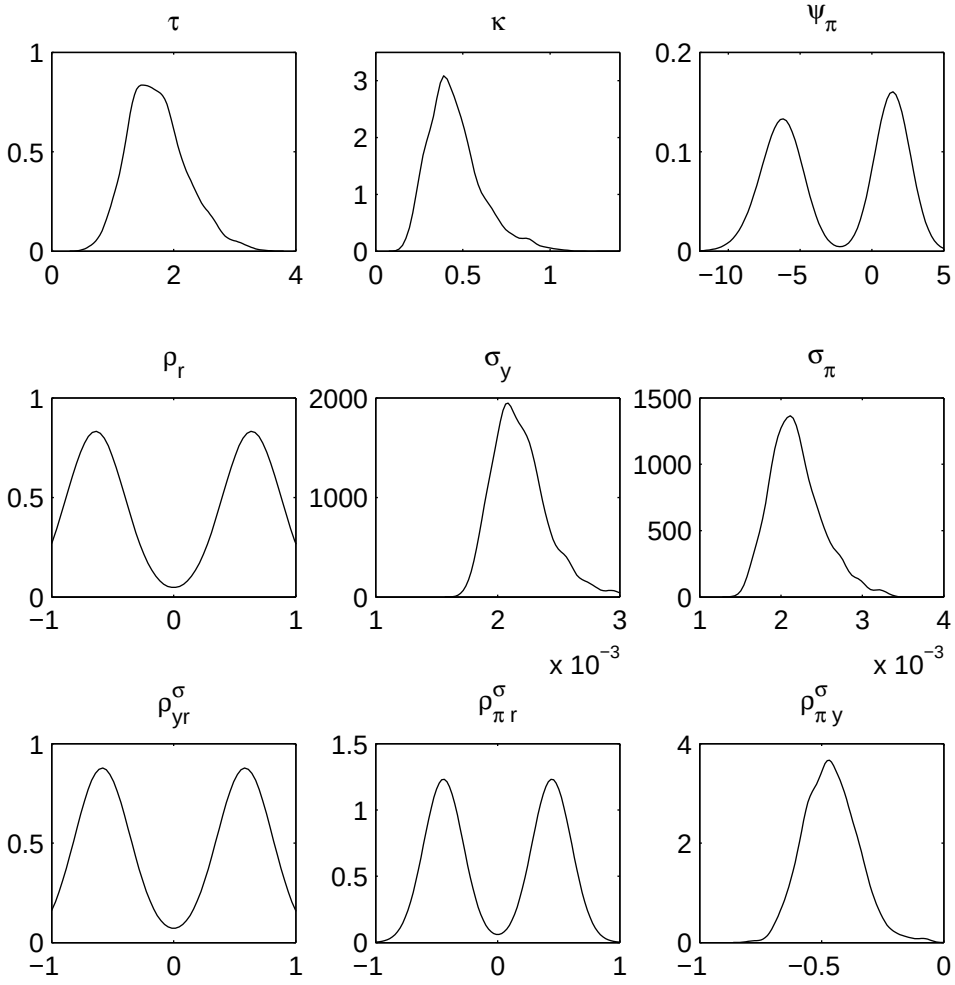
In this section I have expanded on the previous finding of global identifiability problems to demonstrate the implications for the MLE of the CS model. As we have seen previously, the identification pathologies for the AS model are exactly similar. In the next section, I suggest a pragmatic approach to addressing global identification problems in DSGE models which are applicable to all specifications, including those without the analytical solutions and reduced form representations I have exploited thus far. Then, I demonstrate the efficacy of this estimator with respect to both the CS and AS models.

5 MLE using a-priori information

Previously I have designated parameter values for the CS and AS models θ_0, given in Tables 1 and 2 respectively, which are both determinate, and largerly consistent with typically specified priors for DSGE models. These parameters have observationally equivalent indeterminate values θ^*_0, given in Tables 5 and 7 respectively, which are not necessarily compatible with typically specified priors. As I have demonstrated in the last section, this is also reflected in estimation: Under conditional identification Scheme B in Table 9, $\widehat{\theta}_B$ is similar to θ_0 and $\widehat{\theta}^*_B$ is similar to $\widehat{\theta}^*_0$.

When, as now seems likely in general, a given DSGE model is not necessarily globally identifiable on the basis of restrictions implying local identifiability, further identifying restrictions are necessary. These identifying restrictions must partition set-identified points as would otherwise identifiability in order to be useful. One heavy-handed means to this end would be to simply fix ever-more parameters to constants, until conditional global identification of the complement subset is achieved. However, it seems unreasonable to enforce dogmatic certainty over possibly more parameters than is strictly necessary – i.e., those necessary to ensure local identifiability – while allowing the complement subset to go untethered. Furthermore, even if the analyst had fixed all but a few parameters of a given model, she would still not generally be

Figure 4: Distribution of $\{\widehat{\theta}_B, \widehat{\theta}^*_B\}$.



able to prove global identifiability; most DSGE models of empirical signficance do not have either analytical solutions nor reduced form representation, and brute force search may only be applied to bounded parameter spaces.

Instead, let us consider a valuable piece of knowledge gleaned by the previous analysis: Observational equivalence is prevalent, but observationally equivalent points are not always alike, with respect to well-founded prior beliefs. Implicit prior beliefs like these are in fact imposed almost every time a DSGE model is estimated. For instance, numerical search is typically only carried out over a bounded parameter space. The reasons are two-fold. On the one hand, macroeconomists have informative priors about what values DSGE parameters should reasonably have, by virtue of auxiliary microfounded studies. On the other hand, maximum likelihood estimation of DSGE models is numerically challenging. Although there are reliable algorithms available for this purpose (See Andreasen (2010)), the curse of dimensionality restricts the parameter space over which it is reasonable to search. Consider that the well-known Smets and Wouters (2007) model has in excess of 40 estimable parameters. These two factors are in fact the main driving forces behind the popularity of Bayesian estimation of DSGEs.

In order to utilize such a-priori information in the current situation, consider the following augmentation to how likelihood maximization is conducted: In place of specifying a parameter space and finding all observationally equivalent points, let the analyst find only one likelihood-maximizing value which is most-preferred on the basis of a-priori information. That is, in place of a possibly set-valued MLE[18]

$$\Theta_{MLE} = \{\widehat{\theta}_{MLE}, \widehat{\theta}^{*}_{MLE}, \ldots, \widehat{\theta}^{*\ldots*}_{MLE}\} \tag{45}$$

The analyst simply chooses amongst likelihood maximizers the value which maximizes her prior, i.e., using her a-priori information.

$$\widehat{\theta}_{UAP} = \underset{\theta \in \Theta_{MLE}}{\arg\max}\ p(\theta) \tag{46}$$

[18]If the parameters of a given model are not locally identifiable, then continuous ranges of parameter values may maximize the likelihood function. This setup is perfectly amenable to this situation by alternatively defining Θ_{MLE} to be continuous parameter space, or assuming that parameters are fixed to obtain local identifiability before apply the estimator I suggest, as is the procedure used here. Note, if parameters have been fixed to constants before applying the UAP estimator, the implicit assumption is that they have been fixed to population values.

The assumed regularity condition with respect to $p(\theta)$ will be general. Specifically, we only require that the prior ranks the data-generating population value θ_0 as a-priori more likely than any of its observationally equivalent points. We do not require that the prior is entirely correct, meaning that the full distributional profile of the prior need not be correctly specified. This includes a wide range of possible priors, for instance, over individual parameter values, or sign restrictions. The formalization of this requirement is given by Assumption 1.

Assumption 1. *$p(\theta_0) > p(\theta_0^*)$ where θ_0 is the population value of θ and for all $\theta_0^* \neq \theta_0$ such that $\ell(Y;\theta_0^*) = \ell(Y;\theta_0)\ \forall\ Y \in \mathbb{R}^{Tn_Y \times 1}$.*

Note, the UAP estimator is related to the maximum a-posteriori (MAP) estimator, which maximizes the posterior distribution: $\widehat{\theta}_{MAP} = \arg\max_{\theta\in\Theta}\{\ell(\theta|Y) + \ln P(\theta)\}$. A distinction is that the MAP need not equal a likelihood-maximizing value, if priors are informative. The UAP furthermore has two qualities which make it appealing in the current context. First, the estimator is easy to implement, and is in fact sometimes, unknowingly, already in use. For example, consider the analyst who simply estimates a locally or globally unidentified DSGE model without any conditional identification restrictions in place. However, she conducts this numerical search only over a space consistent with her prior beliefs about parameter values. Then, the estimator she is most likely to obtain is the a-priori most likely element of Θ_{MLE}. I will return to this claim momentarily. Second, the estimator in fact retains many qualities of the MLE. For instance, in the null case that θ is actually globally identifiable, then $\Theta_{MLE} = \widehat{\theta}_{MLE}$ so that $\widehat{\theta}_{UAP} = \widehat{\theta}_{MLE}$ by definition. However, even in the case that the MLE is set-valued, the UAP retains the same asymptotic properties of the MLE. This is described by Proposition 4, proven in Appendix A.

Proposition 4. *Under Assumption 1 and the usual regularity assumptions for maximum likelihood estimation excluding identifiability, the using a-priori information maximum likelihood estimator is consistent, asymptotically normal, and efficient.*

The crux of Proposition 4 is that the correctly-specified prior $p(\theta)$ allows the analyst to choose the data-generating value amongst all observationally equivalent points, thereby providing a near-identical stand-in for the traditional assumption of identifiability. Thus, consistency follows by essentially the same proof, and all ensuing proofs

Parameter	Prior	Mean	Std Dev
τ	$\mathcal{G}$	2	0.25
κ	$\mathcal{G}$	0.33	0.25
ψ_π	$\mathcal{G}$	1.5	1
ρ_r	$\mathcal{B}$	0.75	0.2
$100\sigma_y$	$\mathcal{IG}$	0.2	0.1
$100\sigma_\pi$	$\mathcal{IG}$	0.2	0.1
ρ^σ_{yr}	$\mathcal{N}$	0.5	0.25
$\rho^\sigma_{\pi r}$	$\mathcal{N}$	0.5	0.25
$\rho^\sigma_{\pi y}$	$\mathcal{N}$	-0.5	0.25

Table 10: Priors for CS.

of asymptotic normality and efficiency are the same as for the MLE given that consistency. In the next subsection, I substantiate these claims in the context of both the CS model, using typical Bayesian priors for the whole parameter vector.

5.1 CS

Under Scheme B given in Table 9, the parameters β and σ_r are fixed to their population values in θ_0, and the remaining 9 parameters are not globally identifiable. Table 10 provides independent priors for those 9 parameters which are consistent with those utilized in the literatue on Bayesian estimation of DSGEs. The prior $p(\theta)$ is given by the product of these densities.

In Table 11 I repeat the experiment conducted in Table 9, which considered maximum likelihood esimation under Scheme B without global identifiability, but now using the UAP estimator. Since p attributes no mass to $\widehat{\theta}^*_B$, the UAP estimator is defined by exactly the point $\widehat{\theta}_B$. Notice, this prior also has the quality of being a sign restriction, since it favors the point θ_0 to the paradoxical expansionary-constractionary monetary policy point θ^*_0. Furthermore, the ability to discriminate between observationally equivalent points also corrects the distribution of the estimator. It is observed that for each of the three parameters $(\rho_r, \rho^\sigma_{yr}, \rho^\sigma_{\pi r})$, the confidence intervals for the UAP estimator are much closer to what they were in the MLE globally identified case, Scheme A, than Scheme B. Dashes symbolize that confidence intervals are the same in all three cases, and $\times$ indicates that ψ_π is not estimted in Scheme A. Furthermore, the distribution of the estimators has returned to normally-shaped.

The implication of these findings is that identification is not a mandated necessity

Scheme B, using a-priori information.

	τ	κ	ψ_π	ρ_r	$100\sigma_y$	$100\sigma_\pi$	ρ^σ_{yr}	$\rho^\sigma_{\pi r}$	$\rho^\sigma_{\pi y}$
$\widehat{\theta}^{UAP}_B = \widehat{\theta}_B$	1.71	0.43	1.39	0.65	0.21	0.22	0.57	0.45	−0.47
upper(UAP)	2.5	0.66	1.82	0.84	0.25	0.27	0.8	0.56	−0.27
upper(B)	–	–	3.28	0.8	–	–	0.75	0.53	–
upper(A)	–	–	×	0.86	–	–	0.8	0.56	–
lower(A)	–	–	×	0.47	–	–	0.37	0.35	–
lower(B)	–	–	−6.8	−0.8	–	–	−0.75	−0.53	–
lower (UAP)	0.66	0.03	0.73	0.46	0.16	0.14	0.37	0.35	−0.7

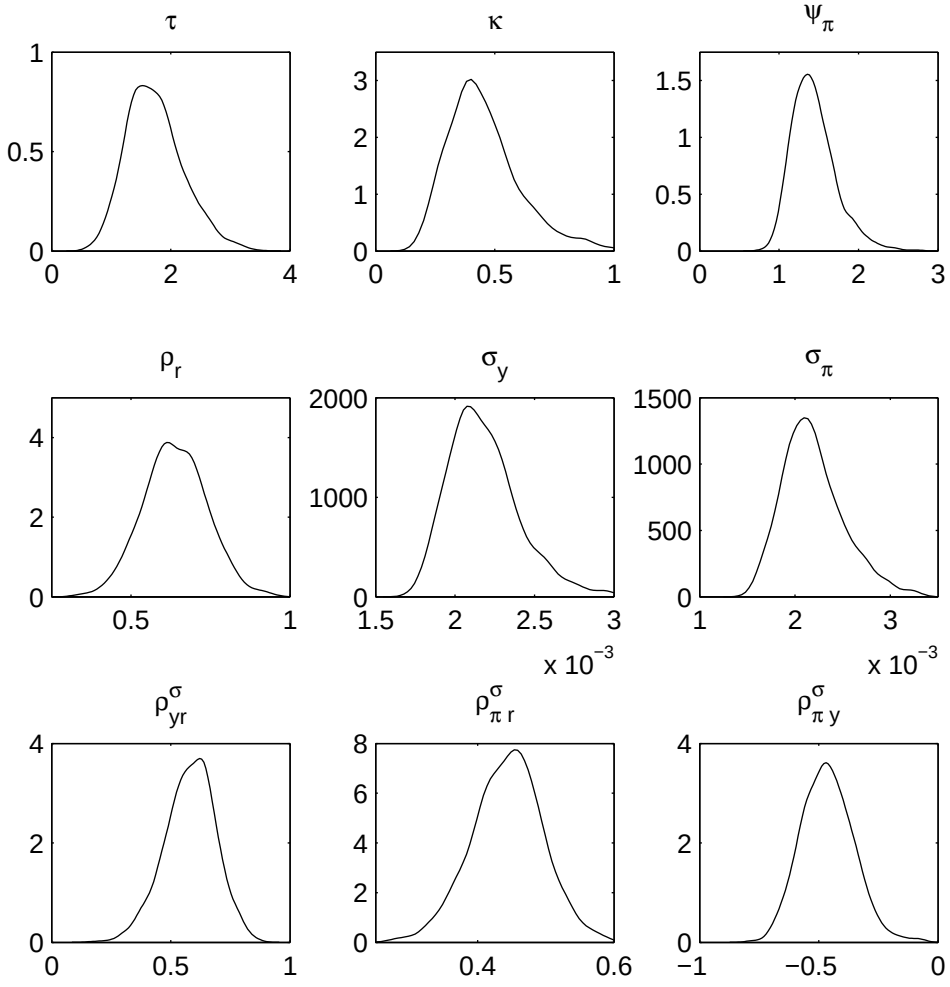


Table 11: Scheme B, using a-priori information. Data-generating θ_0, $T = 250$.

to obtain useful estimators, at least if priors consistent with Assumption 1. This is the case for DSGE models. The intuition is that the analyst need only have some idea of where the estimator should be, and can use that information to sort out observationally equivalent points, but leave sufficient room to let the UAP estimator find the likelihood-maximizing value in that range. When carried out correctly, the small and large sample confidence intervals of the UAP are essentially the same as the MLE, thereby allowing one to surmount the identification challenge. Yet, here the result comes with a caveat; in some situations, one may only necessarily have strong confidence in priors for just a few, or even just one parameter in θ. I now consider how the estimator fares in this situation.

5.2 AS

I now consider the application of the UAP estimator to the AS model. As previously discussed, no conditional identification scheme based on fixing one parameter yields global identifiability. The 5 parameters which differ from point to observationally equivalent point are $(\tau, \rho_z, \sigma_z, \rho^{\sigma}_{gz}, \rho^{\sigma}_{rz})$. Let us consider the situation when the Taylor rule parameter ψ_y is set to its true value in θ_0.

In Table 12, I present the upper and lower bounds for the MLE percentile-bootstrap confidence intervals at θ_0, as well as kernel density approximations of the small-sample distribution for the estimators of ρ_z and ρ^{σ}_{rz}. Once again, the distribution tends to pile up at locations corresponding to the population modes (θ_0, θ_0^*). Furthermore, the MLE confidence intervals tend to be relatively large; see, for instance, a sample estimator from a data set generated by θ_0 in Appendix table C.1.

As remarked previously, the value of the CRRA τ at θ_0, 2, is more in line with common knowledge than the value at θ_0^*, -45.45. Therefore, I now propose to compute the UAP estimator using now just the univariate prior $\tau \sim \mathcal{G}(2, 1)$. Using just this univariate prior is representative of the fact that many macroeconomists have strong convictions about the values of this parameter in particular, and 2 is a commonly argued value. Furthermore, it is possible to use just a prior for this one parameter because all estimates are jointly determined. However, there is evidently no free lunch in using less information to discriminate points in this case. While the secondary nuisance mode in the distribution of ρ_z is mostly eroded, there are evidently many draws which

contribute to a minor secondary mode still for ρ^{σ}_{gz}. Thus, while asymptotically normal, it seems that normality is not necessarily maintained in small samples when priors are not significantly informative to disambiguate between all observationally equivalent points. One should in this case utilize small sample confidence intervals when available, or instead consider utilizing more informative priors.

5.3 General applicability and discussion

I have thus far only explicitly considered models which have analytical solutions and reduced form representation. These features have made it especially straightforward to apply the UAP estimator. However, the estimator is in fact also easy to implement in cases where neither is taken as given, as well. In fact, as I have remarked, the methodology for implementing the UAP estimator is consistent with how analysts typically implement the MLE in practice. Here I clarify this claim. Consider for example a generic univariate log-likelihood function $\ell(x)$. A function which restricts x to be between to arbitrary values a and b is given by

$$r(x) = a - \frac{a-b}{1+\exp\{x\}} \tag{47}$$

Therefore, a pragmatic approach to obtaining the UAP is as follows:

1. Define $z^* = r^{-1}(x^*)$ for x^* the maximizer of $p(x)$.
2. Choose bounds $[b, a]$ such that $p(x) \geq \overline{p} = p(x^*) - \epsilon$ for all $x \in [b, a]$ and small ϵ.
3. Starting from z^*, find $\widehat{z}$ to maximize $\ell(r(z))$ using any numerical algorithm.
4. If the Hessian is negative definite at $\widehat{z}$, stop.
5. If at a corner solution, set $\overline{p} = \overline{p} - \epsilon$, choose new $[a, b]$ accordingly, and try again.
6. The UAP is $\widehat{x} = r(\widehat{z})$ from the last round.

It follows that as the paramer space defined by $r(x)$ is expanded, eventually the analyst will arrive at the likelihood maximizer which maximizes p. In other words, the UAP estimator. This algorithm is in fact not only easy to implement, and easily generalizable to the multivariate case, it is applicable to any numerical maximization algorithm, and very similar to how the maximum likelihood estimator is typically found in practice. Because of numerical difficulties, it is common for analysts to specify a restrictive

	τ	β	κ	ψ_π	ρ_z	ρ_g	ρ_r	$100\sigma_z$	$100\sigma_g$	$100\sigma_r$	ρ^σ_{gz}	ρ^σ_{rz}	ρ^σ_{rg}
upper (UAP)	8.5	–	–	–	0.96	–	–	18.8	–	–	0.92	0.96	–
upper (MLE)	19.8	18.1	6.5	4.2	0.95	0.99	1.3	33.6	19.2	1.3	0.96	0.99	0.5
lower (MLE)	−17.6	−26	−4.7	−3.3	−0.1	0.8	−0.1	0.1	0.5	0.1	−0.99	−0.93	−0.69
lower (UAP)	−8.9	–	–	–	−0.1	–	–	0.06	–	–	−0.99	−0.86	–

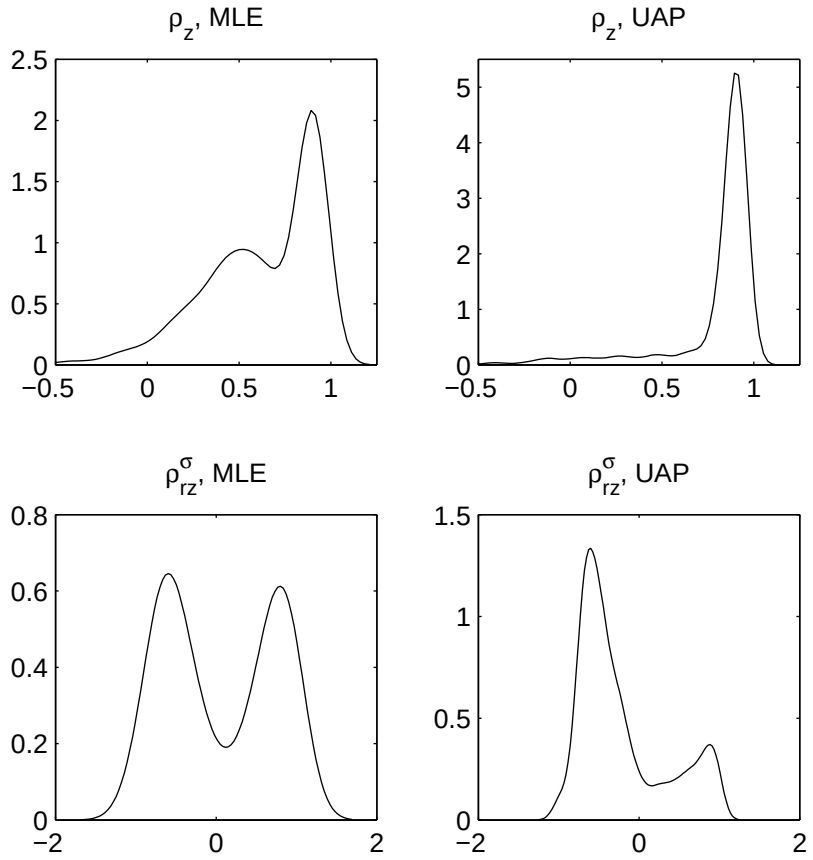


Table 12: 95% percentile bootstrap confidence intervals, θ_0, $T = 250$, $\tau \sim \mathcal{G}(2, 0.25)$ prior.

parameter space over which to find the MLE, and simply expand that space if a corner solution is achieved. What I am suggesting is the same approach.

In some sense the argument developed in this paper is surprising. What I have essentially said is that the analyst who has never even considered global identification may have overlooked observational equivalence, but given she has implemented the MLE by searching over a bounded parameter space consistent with her priors, that may not even be a problem, with respect to asymptotic confidence intervals. However, the conclusion makes sense when one takes stock in just this fact – that the parameter space is consistent with priors – and priors for DSGE models in particular are well-established. It is this dual instance of fates which results in the applicability of the UAP. Yet, an important observation is that the small sample distribution of the estimator will in fact depend on the identification scheme chosen. For this reason, one should compute small sample confidence intervals when possible, as here.

6 Conclusion

This paper has served to achieve three objectives. First, I have documented that identifying restrictions currently in use to achieve local identifiability do not generally achieve global identifiability. This is important since the MLE is not even consistent in this case. Second, I studied this lack of consistency in two cases, and furthermore demonstrated that the distribution of the MLE is consequently non-normal. This is important since currently, non-normality of the MLE in small samples is sometimes attributed to weak identifiability; this is not the case here. Finally, I have offered a modest generalization to the MLE which utilizes readily available priors from macroeconomic theory to address the identification problem, and nests the traditional MLE. This estimator is easy to implement in any DSGE model, not requiring the analyst to find all observationally equivalent points. Furthermore, it is asymptotically efficient. Finally, it is frequently already in use by analysts who maximize the likelihood function only over a parameter space bounded by their priors, known or unknown to them.

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 (2014): "Maximum Likelihood Inference in Weakly Identified DSGE Models," *Forthcoming, 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*, 81.

Bound, J., D. A. Jaeger, and R. Baker (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak," *Journal of the American Statistical Association*, 90, 443–450.

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

Casella, G. and R. L. Berger (2002): *Statistical Inference*, Duxbury Press, 2nd ed.

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. 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.

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.

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

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

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.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton.

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

Hansen, B. E. (2014): *Econometrics*, Unpublished manuscript.

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.

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 (2014): "Global Identificaiton of Linearized DSGE Models," *Working Paper*.

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

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

Lubik, T. and F. Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94, 190–217.

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. (2014): "Inference in DSGE models with possible weak identification," *Quantitative Economics*, 5, 457–494.

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.

——— (2014): "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.

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

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.

——— (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

TAYLOR, J. B. (1993): "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

A Proofs

Proof of Proposition 1. Given $E_t\varepsilon_{t+1} = 0$, the observation equation implies

$$E_t Y_{t+1} = CX_t \tag{A.1}$$

Plugging $E_t y_{t+1}$ and $E_t \pi_{t+1}$, along with $E_t z_{t+1} = \rho_z z_t$, into aggregate demand (21),

$$y_t = \underbrace{\left(c_{yz} + \frac{1}{\tau}c_{\pi z} + \frac{\rho_z}{\tau}\right)}_{f_{yz}} z_t + \underbrace{\left(c_{yg} + \frac{1}{\tau}c_{\pi g} + (1-\rho_g)\right)}_{f_{yg}} g_t + \underbrace{\left(c_{yr} + \frac{1}{\tau}c_{\pi r} - \frac{1}{\tau}\right)}_{f_{yr}} r_t \tag{A.2}$$

Plugging in the expression for $E_t\pi_{t+1}$ from (A.1), along with the expression for y_t just derived in (A.2) into the Phillips curve (22),

$$\begin{aligned}\pi_t = \underbrace{\left(\kappa c_{yz} + \left(\beta + \frac{\kappa}{\tau}\right)c_{\pi z} + \frac{\rho_z \kappa}{\tau}\right)}_{f_{\pi z}} z_t + \underbrace{\left(\kappa c_{yg} + \left(\beta + \frac{\kappa}{\tau}\right)c_{\pi g} - \rho_g \kappa\right)}_{f_{\pi g}} g_t \\ + \underbrace{\left(\kappa c_{yr} + \left(\beta + \frac{\kappa}{\tau}\right)c_{\pi r} - \frac{\kappa}{\tau}\right)}_{f_{\pi r}} r_t\end{aligned} \tag{A.3}$$

Therefore, collecting (A.2) and (A.3), and using the implicitly defined terms $f_{ij}(\theta_s)$,

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ f_{yz} & f_{yg} & f_{yr} \\ f_{\pi z} & f_{\pi g} & f_{\pi r} \end{bmatrix}}_{F(C(\theta_s),\theta_s)} \underbrace{\begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix}}_{X_t} \tag{A.4}$$

Finally, using also the Taylor rule we have the system of implicit functions $g^C(C, \theta_s) = F - MC - m = 0$, written

$$F - \underbrace{\begin{bmatrix} -1 & (1-\rho_r)\psi_y & (1-\rho_r)\psi_\pi \\ 0 & 1 & 1/\tau \\ 0 & \kappa & \beta + \kappa/\tau \end{bmatrix}}_{M(\theta_s)} C - \underbrace{\begin{bmatrix} 0 & -(1-\rho_g)\psi_y\rho_g & 1+\rho_r \\ \rho_z/\tau & 1-\rho_g & -1/\tau \\ \kappa\rho_z/\tau & -\kappa/\rho_g & -\kappa/\tau \end{bmatrix}}_{m(\theta_s)} = 0 \quad \text{(A.5)}$$

g^C makes up the complete system of 9 and equations and 9 unknowns necessary to solve for the elements of C. Although it is infeasible to do this by hand, it is again straightforward to make use of symbolic computation software for this purpose. I do this in two steps (See `solveCompare` in the attached documentation). First, I solve for all 8 elements of C beside c_{rr}. Each of these eight elements has a unique solution in terms of (c_{rr}, θ_s). However, then solving for c_{rr} in terms of only θ_s yields 3 solutions. Therefore, once again we have three solutions for any θ_s. Each has the generalized functional form

$$C^{(i)}(\theta_s) = \begin{bmatrix} c_{rz}^{(i)} & 0 & c_{rr}^{(i)} \\ c_{yz}^{(i)} & \rho_g & c_{yr}^{(i)} \\ c_{\pi z}^{(i)} & 0 & c_{\pi r}^{(i)} \end{bmatrix} \quad \text{(A.6)}$$

Expressions for $\{c_{ij}^{(i)}\}$ in terms of θ_s are too complicated to provide human intuition (See `uap/run/as/solve/solns` in the attached documentation). But, once again, they are known to the computer in closed form.

I now turn to the remaining elements of ABCD representation. Since $Y_t = FX_t$ by (A.4), note that the Taylor rule (20) may be rewritten in the following functional form, for λ_1, λ_2, and λ_3 appropriately defined.

$$\lambda_1 r_t = \rho_r r_{t-1} + \lambda_2 z_t + \lambda_3 g_t + \varepsilon_{rt} \quad \text{(A.7)}$$

Since z_t and g_t are determined by AR(1) processes, it follows immediately that $d_{rz}^{(i)} = c_{rz}^{(i)}/\rho_z$, $d_{rg}^{(i)} = c_{rg}^{(i)}/\rho_g = 0$, and $d_{rr}^{(i)} = c_{rr}^{(i)}/\rho_r$. In other words, $B^{(i)}$ is related to $A^{(i)}$ by

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c_{rz}^{(i)}/\rho_z & 0 & c_{rr}^{(i)}/\rho_r \end{bmatrix}}_{B^{(i)}} \times \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ 0 & 0 & \rho_r \end{bmatrix}}_{\rho(\theta_s)} = \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{\pi z}^{(i)} & 0 & c_{\pi r}^{(i)} \end{bmatrix}}_{A^{(i)}} \quad \text{(A.8)}$$

Furthermore, since $B^{(i)}\rho = A^{(i)}$, we also have $F^{(i)}B^{(i)}\rho = F^{(i)}A^{(i)}$. Note, since $Y_t = FX_t$ and $X_t = AX_{t-1} + B\varepsilon_t$, then $Y_t = FAX_{t-1} + FB\varepsilon_t$ implying $C = FA$ and $D = FB$ exactly. In other words, $D^{(i)}\rho = C^{(i)}$ exactly. Thus, all ABCD matrices may be written exclusively in terms of $C^{(i)}$ and ρ as in (28). □

Proof of Proposition 2. If Y_t has ABCD representation, but also $Y_t = FX_t$ for invertible F, then $FX_t = FAF^{-1}FX_t + FB\varepsilon_t$ and thus $Y_t = FAF^{-1}Y_{t-1} + FB\varepsilon_t$. But also, $C = FA$ and $D = FB$ exactly. Thus, $F = CA^{-1}$ so that $FAF^{-1} = CAC^{-1}$ and $FB = D$. □

Proof of Proposition 3. Let us momentarily drop the (i) superscript for clarity of functional forms. Using the form of the VAR (30) to define $E_tY_{t+1} = \Phi Y_t$, and plugging these expressions into the structural equations (18)-(23) yields

$$\underbrace{\begin{bmatrix} \gamma_0^{(1,1)} & 0 & \gamma_0^{(1,3)} \\ \beta\phi_{\pi r}/\kappa & 1 & (\beta\phi_{\pi\pi}-1)/\kappa \\ \gamma_0^{(3,1)} & 0 & \gamma_0^{(3,3)} \end{bmatrix}}_{\Gamma_0} Y_t = \underbrace{\begin{bmatrix} \rho_z\gamma_0^{(1,1)} & 0 & \rho_z\gamma_0^{(1,3)} \\ \rho_g\beta\phi_{\pi r}/\kappa & \rho_g & \rho_g(\beta\phi_{\pi\pi}-1)/\kappa \\ \rho_r & 0 & 0 \end{bmatrix}}_{\Gamma_1} Y_{t-1} + \varepsilon_t \quad \text{(A.9)}$$

$$\gamma_0^{(1,1)} = \frac{\kappa - \tau\kappa\phi_{yr} - (\kappa + \tau\beta(1-\rho_r))\phi_{\pi r}}{\rho_z\kappa}$$

$$\gamma_0^{(1,3)} = \frac{(1-\rho_g)\tau - \tau\kappa\phi_{yr} - (\kappa + \tau\beta(1-\rho_g))\phi_{\pi\pi}}{\rho_z\kappa}$$

$$\gamma_0^{(3,1)} = 1 + (1-\rho_r)\beta\psi_y/\kappa\phi_{\pi r}$$

$$\gamma_0^{(3,3)} = (1-\rho_r)\beta\psi_y/\kappa\phi_{\pi\pi} - (1-\rho_r)(\psi_\pi + \psi_y/\kappa)$$

These functional forms are each computed symbolically by the function `makeAS` in the attached documentation to ensure no human error. Since now $\Gamma_0 Y_t = \Gamma_1 Y_{t-1} + \varepsilon_t$, but also $Y_t = \Phi Y_{t-1} + D\varepsilon_t$ by Proposition 2 and $D = C\rho^{-1}$ by Proposition 1, then $\Gamma_0 = \rho C^{-1}$. Finally, $\Gamma_0^{-1}\Gamma_1 = \Phi$ implies $\Gamma_1 = \Gamma_0\Phi$. Taking the vec yields $\text{vec}(\Gamma_1) - (I_3 \otimes \Gamma_0)\text{vec}(\Phi)$, and $\text{vec}(\Phi) = R\phi$. □

Proof of Proposition 4. The proof follows from the usual asymptotic properties of the maximum likelihood estimator. The crux of the proposition is that Assumption 1 replaces the usual identifiability regularity condition necessary for consistency, and

asymptotic normality and efficiency follow. To demonstrate consistency, let us assume the typical regularity conditions necessary to prove consistency of the maximum likelihood estimator aside from identifiability hold; see for instance Casella and Berger (2002) p. 516 (A1), (A3), and (A4) for the iid data case. $\widehat{\theta}$ is the maximizer of $p(\theta)$ amongst the possibly many maximizers of $\ell(\theta|Y)$, $\{\widehat{\theta}, \{\widehat{\theta^*}\}\}$. By the weak law of large numbers, for any $\theta \in \Theta$, $T^{-1}\ell(\theta|Y) \xrightarrow{a.s.} \ell(\theta)$ for $\ell(\theta) = E_{\theta_0}\ell(\theta|Y)$ and $\ell(\theta) \leq \ell(\theta_0)$ with equality if and only if $\theta \in \{\theta_0, \{\theta_0^*\}\}$. But $p(\theta_0) > p(\theta_0^*)$ for all $\theta_0^* \neq \theta_0$ by Assumption 1 so that $\widehat{\theta} \xrightarrow{p} \theta_0$. With consistency in hand, let us further assume the typical regularity conditions to prove asymptotic normality and efficiency of the maximum likelihood estimator, i.e. Casella and Berger p. 516 (A5) and (A6). It is then easy to show that $\widehat{\theta} - \theta_0 \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$ using the typical associated proof for the maximum likelihood estimator; this does not further depend on identifiability outside of the previously demonstrated consistency requirement. □

B AS model

The following small scale economy is similar to that presented in An and Schorfheide (2007). A modest generalization is to allow for dependence between shocks. The model consists of a final goods producing firm, a continuum of intermediate goods producing firms, a representative household, and both a monetary and fiscal authority. It abstracts from both wage rigidities and capital accumulation.

B.1 Equilibrium

Final Good Production. A perfectly competitive final goods producing firm has Dixit-Stiglitz type packaging technology, where intermediate goods are numbered by the index of integration j. $1/\nu$ is the elasticity of demand, and the market price for the final good is given by an aggregate P_t of intermediate goods prices $P_t(j)$.

$$Y_t = \left(\int_0^1 Y_t(j)^{1-\nu} \mathrm{d}j \right)^{\frac{1}{1-\nu}} \qquad P_t = \left(\int_0^1 P_t(j)^{\frac{\nu-1}{\nu}} \mathrm{d}j \right)^{\frac{\nu}{\nu-1}}$$

The profit maximization problem is given by the following symmetric maximization for each input good i.

$$\max_{Y_t(i)} P_t \left(\int_0^1 Y_t(j)^{1-\nu} \mathrm{d}j \right)^{\frac{1}{1-\nu}} - \int_0^1 P_t(j) Y_t(j) \mathrm{d}j$$

Since maximization is conducted with respect to specific good i, and j is only an index of integration, differentiation with respect to i and integration with respect to j commute. Thus, the following first order condition.

$$\frac{1}{1-\nu} P_t \left(\int_0^1 Y_t(j)^{1-\nu} \mathrm{d}j \right)^{\frac{\nu}{1-\nu}} (1-\nu) Y_t(i)^{-\nu} - P_t(i) = 0$$

Finally, given that $\left(\int_0^1 Y_t(j)^{1-\nu} \mathrm{d}j \right)^{\nu/(1-\nu)} \equiv Y_t^{\nu}$, profit maximization implies the following demand schedule for intermediate good i.

$$Y_t(i) = \left(\frac{P_t}{P_t(i)} \right)^{1/\nu} Y_t$$

Intermediate Goods Production. Each intermediate good i is produced by intermediate firm i using the following linear technology. A_t is an exogenous productivity process, and $N_t(i)$ is the specific labor input to good i.

$$Y_t(i) = A_t N_t(i)$$

Intermediate firms face nominal rigidities in price adjustment; these are given by the following quadratic costs. ϕ is an index of price stickiness and Π is the steady state inflation rate.

$$\Phi_t(j) = \frac{\phi}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - \Pi \right)^2 Y_t(j) \tag{B.1}$$

Subject to these nominal rigidities, and real wages W_t, each firm i chooses labor and prices to solve the following profit maximization problem. $Q_{t+s|t}$ is the discounted value of future consumption today determined independently by households.

$$\max_{N_t(i), P_t(i)} E_t \left(\sum_{s=0}^{\infty} Q_{t+s|t} \left(\frac{P_{t+s}(i)}{P_{t+s}} Y_{t+s}(i) - W_{t+s} N_{t+s}(i) - \Phi_{t+s}(i) \right) \right)$$

Explicitly plugging in for adjustment costs and output, this problem may also be

written

$$\max_{N_t(i),P_t(i)} E_t \left(\sum_{s=0}^{\infty} Q_{t+s|t} \left(\left(\frac{P_{t+s}(i)}{P_{t+s}} - \frac{\phi}{2} \left(\frac{P_{t+s}(i)}{P_{t+s-1}(i)} - \Pi \right)^2 \right) A_{t+s} - W_{t+s} \right) N_{t+s}(i) \right)$$

Therefore, defining inflation as $\Pi_t = P_t/P_{t-1} = P_t(i)/P_{t-1}(i)$ and recalling the definition of P_t in terms of intermediate good prices given by the Dixit-Stiglitz technology, the first order conditions with respect to $P_t(i)$ and $N_t(i)$ are, respectively,

$$1 - \left(\frac{P_t}{P_t(i)} \right)^{\frac{1-\nu}{\nu}} = \phi \left((\Pi_t - \Pi)\,\Pi_t - E_t \left[Q_{t+1|t} \frac{Y_{t+1}(i)}{Y_t(i)} (\Pi_{t+1} - \Pi)\,\Pi_{t+1} \right] \right) \frac{P_t}{P_t(i)} \quad \text{(B.2)}$$

$$\frac{P_t}{P_t(i)} = \frac{A_t}{W_t + \frac{\phi}{2} A_t (\Pi_t - \Pi)^2} \quad \text{(B.3)}$$

Representative Household. The representative household has real money balances, M_t/P_t, and hours, H_t in the utility function. In addition, consumption provides utility only in proportion to a habit stock, given by the exogenous level of technology, A_t.

$$E_t \left(\sum_{s=0}^{\infty} \beta^s \left(\frac{(C_{t+s}/A_{t+s})^{1-\tau} - 1}{1-\tau} + \chi \ln \left(\frac{M_{t+s}}{P_{t+s}} \right) - H_{t+s} \right) \right)$$

Here, $1/\tau$ is the intertemporal elasticity of substitution (equivalently in this context, the inverse coefficient of relative risk aversion) and χ is a scale factor that determines the steady state of real money balances. The household may trade bonds B_t at gross nominal rate R_t, pays lump-sum taxes T_t, and receives a net cash inflow from trading a set of state-contingent securities X_t. Given these features, the household's budget constraint is

$$P_t C_t + B_t + P_t \sum_{S_{t+1}} Q_{t+1|t} X_{t+1} + M_t + T_t = P_t W_t H_t + R_{t-1} B_{t-1} + P_t X_t + M_{t-1}$$

where S_{t+1} is the realization of the state in period $t+1$. The first order conditions of the corresponding Lagrangean (multiplier λ_t) with respect to X_{t+1}, C_t, and B_t are, respectively,

$$\beta E_t \frac{\lambda_{t+1}}{\lambda_t} = \frac{1}{\Pi_{t+1}} Q_{t+1|t} \quad \text{(B.4)}$$

$$\beta E_t \frac{\lambda_{t+1}}{\lambda_t} = \beta E_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{A_t}{A_{t+1}} \frac{1}{\Pi_{t+1}} \right] \quad \text{(B.5)}$$

$$\beta E_t \frac{\lambda_{t+1}}{\lambda_t} = \frac{1}{R_t} \tag{B.6}$$

Finally, combining the first order conditions with respect to C_t and H_t yields the following expression; the first order condition with respect to M_t is not stated since it will not be of use in the log-linearized solution.

$$W_t = A_t (C_t/A_t)^{\tau} \tag{B.7}$$

Partial Equilibrium Between Firms and Households. Firstly, plugging Equation (B.7) into (B.3) gives

$$\frac{P_t}{P_t(i)} = \frac{1}{(C_t/A_t)^{\tau} + \phi/2(\Pi_t - \Pi)^2}$$

while combining Equations (B.4) and (B.5) gives

$$Q_{t+1|t} = \beta E_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{A_t}{A_{t+1}} \right]$$

Inputting these last two equalities into Equation (B.2), and using the steady state-local approximation $(P_t/P_t(i))^{\frac{1-\nu}{\nu}} \approx \frac{1-\nu}{\nu}(P_t/P_t(i) - 1)$ gives

$$\begin{aligned} 1 = \frac{1}{\nu}\left(1 - \left(\frac{C_t}{A_t}\right)^{\tau}\right) + \phi(\Pi_t - \Pi)\left(\left(1 - \frac{1}{2\nu}\right)\Pi_t + \frac{\Pi}{2\nu}\right) \\ - \phi\beta E_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{Y_{t+1}/A_{t+1}}{Y_t/A_t} (\Pi_{t+1} - \Pi)\Pi_{t+1} \right] \end{aligned} \tag{B.8}$$

Secondly, combining the household first order conditions with respect to C_t and B_t, Equations (B.5) and (B.6), yields

$$1 = \beta E_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{A_t}{A_{t+1}} \frac{R_t}{\Pi_{t+1}} \right] \tag{B.9}$$

Exogenous Processes and Market Clearing. The Taylor rule considered in this paper is an extension of Taylor (1993)'s original specification, allowing for lagged interest rates to enter into the monetary authority's decision making process. It is written as follows, where R_t^* denotes the target gross nominal rate and ε_{rt} is an idiosyncratic monetary policy shock:

$$R_t = R_{t-1}^{\rho_r} R_t^{*1-\rho_r} \exp(\varepsilon_{rt}) \tag{B.10}$$

$$R_t^* = \mu\Pi \left(\Pi_t/\Pi\right)^{\psi_\pi} \left(Y_t/Y_t^*\right)^{\psi_y} \tag{B.11}$$

μ is the steady state of real gross interest rates, R_t/Π_t. Π, the steady state of inflation, is also the inflation target. Finally, Y_t^* is the level of output that would prevail without price rigidities ($\phi = 0$), sometimes also known as the "natural" rate.

The fiscal authority consumes a portion of output, $F_t = \zeta_t Y_t$, and levies a lump sum tax, T_t, subject to a budget constraint

$$P_t F_t + R_{t-1} B_{t-1} + M_{t-1} = T_t + B_t + M_t$$

where ζ_t a nonstandard exogenous process which amounts to a function of an AR(1) process. Specifically, defining $G_t = 1/(1-\zeta_t)$ (i.e. $\zeta_t = (G_t - 1)/G_t$), then

$$\ln G_t = (1-\rho_g) \ln G + \rho_g \ln G_{t-1} + \varepsilon_{gt} \tag{B.12}$$

Meanwhile, aggregate productivity follows

$$\ln A_t = \ln \gamma + \ln A_{t-1} + \ln Z_t \tag{B.13}$$

$$\ln Z_t = \rho_z \ln Z_{t-1} + \varepsilon_{zt} \tag{B.14}$$

and γ is the average growth rate of productivity. Finally, market clearing is given by $H_t = N_t$ and the aggregate accounting equality

$$Y_t = C_t + F_t + \Phi_t \tag{B.15}$$

where Φ_t is the adjustment cost in the symmetric equilibrium case. In more straightforward terms, symmetry implies the dependence of $\Phi_t(j)$ on j in Equation (B.1) does not matter, so that Φ_t may be written

$$\Phi_t = \frac{\phi}{2} \left(\Pi_t - \Pi\right)^2 Y_t$$

Thus, using this definition of Φ_t and $F_t = \zeta_t Y_t$, the aggregate accounting equality

Equation (B.15) may also be written

$$Y_t = \frac{C_t G_t}{1 - \frac{\phi}{2}\left(\Pi_t - \Pi\right)^2 G_t} \tag{B.16}$$

Shocks. There are three shocks – to TFP, government spending, and interest rates – respectively ε_{zt}, ε_{gt}, and ε_{rt}. These are dependently mean-zero normally distributed with standard deviations σ_z, σ_g, and σ_r, respectively, and covariances σ_{gz}, σ_{rz}, and σ_{rg}. Correlations are defined by $\rho^{\sigma}_{gz} = \sigma_{gz}/(\sigma_g\sigma_z)$, $\rho^{\sigma}_{rz} = \sigma_{rz}/(\sigma_r\sigma_z)$, and $\rho^{\sigma}_{rg} = \sigma_{rg}/(\sigma_r\sigma_g)$.

Nonlinear Equlibrium Equations and Solution. Simply rewriting (B.8) gives

$$\begin{aligned}\frac{1}{\nu}\left(1 - \exp\left\{\tau \ln \widehat{C}_t\right\}\right) + \phi(\exp\left\{\ln \Pi_t\right\} - \Pi)\left(\left(1 - \frac{1}{2\nu}\right)\exp\left\{\ln \Pi_t\right\} + \frac{\Pi}{2\nu}\right) - \\ \phi\beta E_t\left[\exp\left\{\ln(\widehat{Y}_{t+1}) - \ln(\widehat{Y}_t) - \tau\left(\ln(\widehat{C}_{t+1}) - \ln(\widehat{C}_t)\right)\right\} \times \right. \\ \left. (\exp\left\{\ln \Pi_{t+1}\right\} - \Pi)\exp\left\{\ln \Pi_t\right\}\right] - 1 = 0\end{aligned} \tag{B.17}$$

Meanwhile, defining $\hat{C}_t = C_t/A_t$ and using the equality $\ln A_t - \ln A_{t+1} = -\ln\gamma - \ln Z_{t+1}$ from Equation (B.13), Equation (B.9) may be written

$$\beta E_t\left[\exp\left\{\ln R_t - \ln \Pi_{t+1} - \tau\left(\ln(\widehat{C}_{t+1}) - \ln(\widehat{C}_t)\right) - \ln\gamma - \ln Z_{t+1}\right\}\right] - 1 = 0 \tag{B.18}$$

Equation (B.10) is simply rewritten

$$\ln R_t + \rho_r \ln R_{t-1} + (1 - \rho_r)\ln R^*_t + \varepsilon_{rt}$$

R^*_t, $\widehat{Y}^*_t = Y^*_t/A_t$, and $\widehat{C}^*_t = C^*_t/A_t$ are defined to exist in a world exactly like the model economy, but with $\phi = 0$. When $\phi = 0$, Equation (B.15) implies $\widehat{Y}^*_t = G_t\widehat{C}^*_t$. But also, when $\phi = 0$, Equation (B.8) implies $\widehat{C}^*_t = (1-\nu)^{1/\tau}$. Therefore, $\ln \widehat{Y}^*_t = (1/\tau)\ln(1-\nu) + \ln G_t$. Using these facts, Equation (B.11) may be rewritten as

$$\ln R^*_t = \ln\mu + (1 - \psi_\pi)\ln\Pi + \psi_\pi \ln \Pi_t + \psi_y(\ln \widehat{Y}_t - \ln G_t - (1/\tau)\ln(1-\nu))$$

The last two equations above may simply be combined to eliminate the variable R_t^*:

$$-\ln R_t + (1-\rho_r)\left(\ln\mu + (1-\psi_\pi)\ln\Pi - \psi_y(1/\tau)\ln(1-\nu)\right) + \rho_r \ln R_{t-1} + (1-\rho_r)\psi_\pi \ln\Pi_t + (1-\rho_r)\psi_y(\ln\widehat{Y}_t - \ln G_t) + \varepsilon_{rt} = 0 \quad \text{(B.19)}$$

Dividing both sides of Equation (B.16) through by A_t and rewriting yields

$$-\exp\left\{\ln\widehat{Y}_t\right\} + \exp\left\{\ln\widehat{C}_t + \ln G_t\right\} \Big/ \left(1 - \frac{\phi}{2}(\exp\{\ln\Pi_t\} - \Pi)^2 \exp\{\ln G_t\}\right) = 0 \quad \text{(B.20)}$$

Finally, Equations (B.14) and (B.12) simply restated are

$$-\ln Z_t + \rho_z \ln Z_{t-1} + \varepsilon_{zt} = 0 \quad \text{(B.21)}$$

$$-\ln G_t + (1-\rho_g)\ln G + \rho_g \ln G_{t-1} + \varepsilon_{gt} = 0 \quad \text{(B.22)}$$

In all, there are six equilibrium equations (B.17) - (B.22) which characterize equilibrium for this model stated in terms of the six (detrended) variables of interest $\ln Z_t$, $\ln G_t$, $\ln R_t$, $\ln\hat{Y}_t = \ln(Y_t/A_t)$, $\ln\Pi_t$, and $\ln\hat{C}_t = \ln(C_t/A_t)$. The shocks are ε_{zt}, ε_{gt}, and ε_{rt}. The linearized versions of these model are given in (18)-(23).

Steady State. The equilibrium equations also characterize steady state. Equation (B.18) implies the steady state of nominal gross interest rates is $R = \gamma\Pi/\beta$ because Equation (B.21) implies $Z = 1$. Furthermore, $\mu = R/\Pi$ by definition so $\mu = \gamma/\beta$. Equation (B.17) implies $\widehat{C} = (1-\nu)^{1/\tau}$ because $\Pi_t = \Pi$ by definition in the steady state. Given this, Equation (B.20) implies $\widehat{Y} = G(1-\nu)^{1/\tau}$. The steady states Π and G are structural parameters.

Linearization. Linearization of (B.17) - (B.22) yields (18)-(23) in the text. A summary of all structural parameters and variables in the model is given in Table (B.1). Lower-case variables denote log-deviations from steady state, for instance $\pi_t = \ln\Pi_t - \ln\Pi$. Note, the parameters γ and G are unidentifiable from (18)-(23). This latter point is discussed in the next subsection.

	Structural Params (18)		**Endogenous** (6)		**Shocks** (3)	
1	τ	CRRA	z_t	Total factor prod.	ε_{zt}	TFP
2	β	Discount factor	g_t	Gov spending	ε_{gt}	Gov sp.
3	ν	Inverse elas. of demnd	r_t	Nominal int rate	ε_{rt}	Nom. int.
4	ϕ	Index of price stckness	y_t	Nominal output		
5	γ	Avg. gr. rate of prod.	π_t	Inflation		
6	Π	St. state level of infl.	c_t	Nominal cons.		
7	G	St. state level of G_t.				
8	ψ_π	Taylor rule infl. coeff.				
9	ψ_y	Taylor rule out. coeff.				
10	ρ_z	z_t persistence				
11	ρ_g	g_t persistence				
12	ρ_r	r_t persistence				
13	σ_z	ε_{zt} std error				
14	σ_g	ε_{gt} std error				
15	σ_r	ε_{rt} std error				
16	σ_{gz}	Cov of ε_{gt} and ε_{zt}				
17	σ_{rz}	Cov of ε_{rt} and ε_{zt}				
18	σ_{rg}	Cov of ε_{rt} and ε_{gt}				

Table B.1: AS model parameter and variable names.

B.2 AS with-means

The body of the paper considers only the case of data that has been differenced from means. Let us consider also the case when means of the data are used. As shown above, the observables in VAR representation (30) are differences from steady states; $r_t = \ln R_t - \ln R$, $y_t = \ln Y_t - \ln Y$, and $\pi_t = \ln \Pi_t - \ln \Pi$. Provided that the unconditional means of the data are used as a proxy for steady state, we also have the following rule of motion for $V_t = \begin{bmatrix} \ln R_t & \ln \widehat{Y}_t & \ln \Pi_t \end{bmatrix}'$.

$$V_t = \psi(\theta_s^\psi) + \Phi^{(i)}(\theta_s^\psi) V_{t-1} + U_t^{(i)} \tag{B.23}$$

where

$$\underbrace{\begin{bmatrix} \psi_r^{(i)} & \psi_y^{(i)} & \psi_\pi^{(i)} \end{bmatrix}'}_{\psi^{(i)}(\theta_s^\psi)} = (I_3 - \Phi^{(i)}) \underbrace{\ln \left[\begin{array}{c:c:c} \gamma\Pi/\beta & G(1-\nu)^{1/\tau} & \Pi \end{array} \right]'}_{V(\theta_s^\psi)}$$

for $V = \ln \begin{bmatrix} R & \widehat{Y} & \widehat{\Pi} \end{bmatrix}'$ and

$$\underset{12\times 1}{\theta_s^\psi} = (\gamma, G, \rho_g, \beta, \nu, \phi, \Pi, \tau, \rho_z, \psi_y, \psi_\pi, \rho_r)' \tag{B.24}$$

Note, θ_s^{ψ} is similar to θ_s, with two exceptions. First, both γ and G, the average growth rate of productivity and steady state government spending, now appear. These are entirely unidentifiable using differenced-from-means data Y_t, but appear in $\psi^{(i)}$. In addition, κ is replaced with (ν, ϕ, Π). This is due to the fact that the elements of $\psi^{(i)}$ possibly relinquish the previous linear dependence between these variables in κ. Finally, we may define the 3×1 implicit function

$$g^{\psi}(\psi^{(i)}, \phi^{(i)}, \theta_s^{\psi}) = \exp\{V\} - \exp\{(I_3 - \Phi^{(i)})^{-1}\psi^{(i)}\} = 0 \tag{B.25}$$

to construct the extended 16×1 set of moment conditions $g^{\psi}(\pi^{\psi(i)}, \theta^{\psi}) = 0$, written

$$\begin{bmatrix} g^{\psi}(\psi^{(i)}, \phi^{(i)}, \theta_s^{\psi}) \\ g^{\phi}(\phi^{(i)}, \theta_s^{\psi}) \\ g^{\omega}(\pi^{(i)}, \theta^{\psi}) \end{bmatrix} = \begin{bmatrix} 0_{3\times 1} \\ 0_{7\times 1} \\ 0_{6\times 1} \end{bmatrix} \tag{B.26}$$

For structural and reduced from parameters

$$\underset{18\times 1}{\theta^{\psi}} = (\theta_s^{\psi\prime}, \theta_{\sigma}')' \tag{B.27}$$

$$\underset{16\times 1}{\pi^{(i)\psi}} = (\psi^{(i)\prime}, \phi^{(i)\prime}, \omega^{(i)\prime})' \tag{B.28}$$

B.2.1 AS with-means identifiability

Using the above setup, in Table B.2 I consider the identifiability of the AS model with-means. Since there are 18 reduced form parameters and 16 structural parameters, we fix two parameters to constants. As in the case with-means, all parameters of the Taylor rule are not identifiable. Thus, one of $(\psi_y, \psi_\pi, \rho_r)$ or $(\sigma_{rz}, \sigma_{rg}, \sigma_r^2)$ must be fixed for even local identifiability. Second, one of the 6 parameters G, γ, β, ν, ϕ, or Π must be fixed for local identifiability. This is because there are only 5 degrees of freedom in $(g^{\psi}, g_2^{\phi}, g_6^{\phi})$. Thus, neither τ nor ρ_z is fixed and thus the 2 in row (g_1^{ϕ}, g_5^{ϕ}) implies any scheme implying local identifiability by fixing two parameters will not imply global.

		θ_s^ψ									θ_σ						
		γ	G	ρ_g	β	ν	ϕ	Π	(τ, ρ_z)	$(\psi_y, \psi_\pi, \rho_r)$	σ_z	σ_{gz}	σ_{rz}	σ_g^2	σ_{rg}	σ_r^2
g^ψ	g_1^ψ	✓			✓			✓									
	g_2^ψ		✓			✓			[✓,]								
	g_3^ψ							✓									
g^ϕ	g_4^ϕ			✓													
	(g_2^ϕ, g_6^ϕ)			$✓_{2\times1}$	✓	$✓_{2\times1}$	$✓_{2\times1}$	$✓_{2\times1}$									
	(g_1^ϕ, g_5^ϕ)			$✓_{2\times1}$	✓	$✓_{2\times1}$	$✓_{2\times1}$	$✓_{2\times1}$	2								
	(g_3^ϕ, g_7^ϕ)				✓					$✓_{2\times3}$							
g^ω	g_1^ω			2	$2_{1\times2}$				$2_{1\times2}$		✓						
	g_2^ω			✓	$2_{1\times2}$				$✓_{1\times2}$			✓					
	g_3^ω			✓	$2_{1\times2}$				$✓_{1\times2}$	$✓_{1\times3}$			✓				
	g_4^ω				$2_{1\times2}$									✓			
	g_5^ω				$2_{1\times2}$					$✓_{1\times3}$					✓		
	g_6^ω				$2_{1\times2}$					$2_{1\times3}$						✓	

Table B.2: Identifiability in AS model on the basis of g: With-means.

C Estimation

C.1 FGLS estimation of reduced form parameters

This section details the FGLS estimator of the reduced form parameters $\widehat{\pi} = \left(\widehat{\phi}', \widehat{\omega}'\right)'$. Consider the process

$$Y_t = \Phi Y_{t-1} + U_t \tag{C.1}$$

U_t is NID mean zero covariance Ω. Since Ω is positive semidefinite, it has Cholesky decomposition $\Omega = LL'$ for L a lower-diagonal matrix. Since L is lower-diagonal, it is full rank. So,

$$L^{-1}Y_t = L^{-1}\Phi Y_{t-1} + L^{-1}U_t$$

where $L^{-1}U_t$ has covariance matrix I. Thus, we have

$$L^{-1}E\left(Y_t Y_{t-1}'\right)L^{'-1} = L^{-1}\Phi E\left(Y_{t-1}Y_{t-1}'\right)L^{'-1}$$

Assume data is available for time sample $t = 1, \ldots, T$. This suggests the estimator $\widehat{\Phi}$,

$$L^{-1}\frac{1}{T-1}\sum_{t=2}^{T} Y_t Y_{t-1}' L^{'-1} = L^{-1}\widehat{\Phi}\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}Y_{t-1}' L^{'-1} \tag{C.2}$$

In the models considered in the paper, Φ contains exclusion restrictions. Let the nonzero elements of Φ be collected in the vector ϕ, which is related to Φ by the equation $\text{vec}(\Phi) = R\phi$. See (33). The matrix R is an appropriately defined zero-one selection matrix R. Vectorizing (C.2) therefore yields

$$(L\otimes L)^{-1}\text{vec}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t Y_{t-1}'\right) = (L\otimes L)^{-1}\left(\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}Y_{t-1}'\right)\otimes I_3\right)R\widehat{\phi}$$

If Ω were known, then the GLS estimator for ϕ would therefore be written as follows; this follows by simply premultiplying the last equality by R and inverting:

$$\widehat{\phi} = \left(R'(L\otimes L)^{-1}\left(\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}Y_{t-1}'\right)\otimes I\right)R\right)^{-1} \times R'(L\otimes L)^{-1}\text{vec}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t Y_{t-1}'\right) \tag{C.3}$$

Here, with Ω also unknown, we require joint feasible generalized least squares (FGLS) estimation of ϕ and Ω. Pseudocode for an iterated FGLS estimator follows.

1. Guess $\Omega = I$. Compute the consistent, albeit inefficient estimator

$$\widehat{\phi}_* = \left(R' \left(\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_{t-1} Y_{t-1}' \right) \otimes I \right) R \right)^{-1} \times R' \text{vec} \left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t Y_{t-1}' \right)$$

2. Given $\widehat{\phi}_*$ define the fitted values $\widehat{U}_t = Y_t - (Y_{t-1}' \otimes I) R \widehat{\phi}_*$. Compute

$$\widehat{L}_* = \text{chol} \left(\frac{1}{T-1} \sum_{t=2}^{T} \widehat{U}_t \widehat{U}_t' \right)$$

3. Given $\widehat{L}_*$ compute the FGLS estimator $\widehat{\phi}$.

$$\widehat{\phi} = \left(R' (\widehat{L}_* \otimes \widehat{L}_*)^{-1} \left(\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_{t-1} Y_{t-1}' \right) \otimes I \right) R \right)^{-1} \times R' (\widehat{L}_* \otimes \widehat{L}_*)^{-1} \text{vec} \left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t Y_{t-1}' \right) \quad \text{(C.4)}$$

4. Calculate the FGLS estimator $\widehat{L}$ given $\widehat{\phi}$ similarly to Step 2.

5. If $\widehat{\phi}$ and $\widehat{L}$ are sufficiently close to $\widehat{\phi}_*$ and $\widehat{L}_*$, stop. Otherwise, iterate until so.

6. Define $\widehat{\omega} = \text{vech}(\widehat{\Omega})$ and $\widehat{\pi} = \left(\widehat{\phi}', \widehat{\omega}' \right)'$.

C.2 Small sample confidence intervals

To compute small sample standard errors by bootstrap, first estimate $\widehat{\pi}$, and obtain the fitted values $\{\widehat{U}_t\}_{t=1}^T$

$$\widehat{U}_t = Y_t - (Y_{t-1}' \otimes I) R \widehat{\phi} \quad \text{(C.5)}$$

Assuming $Y_0 = 0$. Then, draw with replacement from $\{\widehat{U}_t\}_{t=1}^T$ a total of $B = 1000$ times, using these draws and $\widehat{\pi}$ to generate B synthetic data series $\{\{Y_t^{(b)}\}_{t=1}^T\}_{b=1}^B$. Using the FGLS estimator on each of these series, obtain $\{\widehat{\pi}^{(b)}\}$. Then, using the mapping g, obtain the set of draws $\{\widehat{\theta}^{(b)}\}_{b=1}^B$ from $\widehat{\theta}$'s distribution. When there are two values $\{\widehat{\theta}^{(b)}, \widehat{\theta}^{(b)*}\}$ corresponding to a given $\widehat{\pi}^{(b)}$, their weighting depends on which estimator is being utilized. For the MLE, they are weighted equally. If the UAP is used, then the prior-favored point receives a weight of two, and the other zero. When

the UAP is being utilized and for one reason or another the prior does not differentiate between the two points in a given sample, they receive MLE weights. Then, for the k-dimensional vector θ, denote each scalar element θ_i for $i = 1, \ldots, k$, and for each θ_i define a set of deviations

$$\mathbb{S}_i = \left\{ \widehat{\theta}_i^{(1)} - \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \widehat{\theta}_i^{(b)}, \ldots, \widehat{\theta}_i^{(\mathcal{B})} - \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \widehat{\theta}_i^{(b)} \right\} \tag{C.6}$$

Where $\mathcal{B} \geq B$ is the number of draws $\{\widehat{\theta}^{(b)}\}$ which accumulate from the B reduced form draws $\{\widehat{\pi}^{(b)}\}$ due to observational equivalence in each draw. Following Hansen (2014) p. 233, where q_i is the quantile function of $\mathbb{S}_i$, the $1 - \alpha$ bootstrap confidence interval of $\widehat{\theta}_i^*$ is

$$C_{1-\alpha}(\theta_i) = \left[\widehat{\theta}_i - q_i(1 - \alpha/2), \quad \widehat{\theta}_i - q_i(\alpha/2) \right] \tag{C.7}$$

$C_{1-\alpha}$ requires no assumptions about U_t to be correctly sized.

C.3 Supplementary estimated statistics

	$\boldsymbol{\tau}$	β	κ	ψ_π	$\boldsymbol{\rho_z}$	ρ_g	ρ_r	$\mathbf{100}\boldsymbol{\sigma_z}$	$100\sigma_g$	$100\sigma_r$	$\boldsymbol{\rho}^{\boldsymbol{\sigma}}_{\boldsymbol{gz}}$	$\boldsymbol{\rho}^{\boldsymbol{\sigma}}_{\boldsymbol{rz}}$	ρ^{σ}_{rg}
$\widehat{\theta}$	0.6	−7.8	2.4	1.4	0.9	0.98	0.5	0.2	0.6	0.4	−0.1	−0.7	−0.3
$\widehat{\theta}^*$	−3.4	−7.8	2.4	1.4	0.16	0.98	0.5	9.1	0.6	0.4	−0.18	0.97	−0.3

Solution i	$\Phi^{(i)}$	1e5 $\times$ $\Omega^{(i)}$	λ	Stable?
1	$\begin{bmatrix} -0.35 & 0 & 2.56 \\ -3.07 & 0.98 & 5.99 \\ -0.75 & 0 & 2.44 \end{bmatrix}$	$\begin{bmatrix} 8.7 & \cdot & \cdot \\ 22.8 & 65 & \cdot \\ 5.4 & 14.3 & 3.4 \end{bmatrix}$	$\begin{bmatrix} 1.18 \\ 0.98 \\ 0.90 \end{bmatrix}$	no
2	$\begin{bmatrix} 0.80 & 0 & 0.22 \\ 0.22 & 0.98 & -0.70 \\ 0.31 & 0 & 0.26 \end{bmatrix}$	$\begin{bmatrix} 2.8 & \cdot & \cdot \\ 6.7 & 11.8 & \cdot \\ 0.5 & 5.1 & 3.2 \end{bmatrix}$	$\begin{bmatrix} 0.98 \\ 0.9 \\ 0.16 \end{bmatrix}$	yes
3	$\begin{bmatrix} 1.08 & 0 & -0.37 \\ -0.69 & 0.98 & 1.44 \\ 0.70 & 0 & -0.53 \end{bmatrix}$	$\begin{bmatrix} 3.4 & \cdot & \cdot \\ -2.5 & 5.4 & \cdot \\ 7.9 & 6.2 & 19.5 \end{bmatrix}$	$\begin{bmatrix} 0.98 \\ 0.9 \\ -0.35 \end{bmatrix}$	yes

Table C.1: Estimators for AS model. Data-generating θ_0, $T = 250$.

Scheme A: Fix $\boldsymbol{\beta}$ and $\boldsymbol{\psi_\pi}$ to their true values in $\boldsymbol{\theta_0}$.

Reduced form parameters at $\widehat{\theta}_A$.

Solution i	$\begin{bmatrix}\phi^{(i)} & 0\end{bmatrix}$	$1e5 \times \Omega^{(i)}$	Stable?
1	$\begin{bmatrix} 0.41 & 0 & 0 \\ -0.57 & 0 & 0 \\ -0.42 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.16 & \cdot & \cdot \\ -0.07 & 0.34 & \cdot \\ 0.03 & -0.22 & 0.14 \end{bmatrix}$	yes
2	$\begin{bmatrix} 1.06+0.7i & 0 & 0 \\ 0.26-0.6i & 0 & 0 \\ 0.38+0.2i & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.56 & \cdot & \cdot \\ 0.23-1.1i & 1.11 & \cdot \\ 1.01-0.38i & 0.2+0.65i & 0.88 \end{bmatrix}$	no
3	$\begin{bmatrix} 1.06-0.7i & 0 & 0 \\ 0.26+0.6i & 0 & 0 \\ 0.38-0.2i & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1.56 & \cdot & \cdot \\ 0.23+1.1i & 1.10 & \cdot \\ 1.01+0.4i & 0.2-0.65i & 0.88 \end{bmatrix}$	no

Scheme B: Fix $\boldsymbol{\beta}$ and $\boldsymbol{\sigma_r}$ to their true values in $\boldsymbol{\theta_0}$.

Reduced form parameters at $\widehat{\theta}^*_B$. At $\widehat{\theta}_B$ same as $\widehat{\theta}_A$ above.

Solution i	$\begin{bmatrix}\phi^{(i)} & 0\end{bmatrix}$	$1e5 \times \Omega^{(i)}$	Stable?
1	$\begin{bmatrix} 3.25 & 0 & 0 \\ 1.01 & 0 & 0 \\ -0.19 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 10.1 & \cdot & \cdot \\ 4.36 & 2.20 & \cdot \\ 0.9 & 0.2 & 0.2 \end{bmatrix}$	no
2	$\begin{bmatrix} -0.49 & 0 & 0 \\ 0.18 & 0 & 0 \\ 0.05 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.23 & \cdot & \cdot \\ -0.27 & 0.63 & \cdot \\ -0.25 & 0.09 & 0.41 \end{bmatrix}$	yes
3	$\begin{bmatrix} 0.41 & 0 & 0 \\ -0.57 & 0 & 0 \\ -0.42 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.16 & \cdot & \cdot \\ -0.07 & 0.34 & \cdot \\ 0.03 & -0.22 & 0.14 \end{bmatrix}$	yes

Table C.2: Estimators of CS reduced form parameters.