UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Identification of Dynamic Stochastic General Equilibrium Models

Permalink

https://escholarship.org/uc/item/0nc5n2v5

Author

Morris, Stephen David

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Identification of Dynamic Stochastic General Equilibrium Models

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Stephen David Morris

Committee in charge:

Professor James D. Hamilton, Chair
Professor Thomas Barranga
Professor Ivana Komunjer
Professor Valerie Ramey
Professor Rossen Valkanov

2014

The dissertation of Stephen David Morris is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2014

TABLE OF CONTENTS

VITA

2007	B.A. in Economics and Mathematics, Wesleyan University
2014	Ph.D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Identification of Dynamic Stochastic General Equilibrium Models

by

Stephen David Morris

Doctor of Philosophy in Economics

University of California, San Diego, 2014

Professor James D. Hamilton, Chair

The dissertation "Identification of Dynamic Stochastic General Equilibrium Models" by Stephen David Morris is divided into three chapters. The first chapter considers the statistical implications of common identifying restrictions for DSGE models. The second chapter considers the implications of identification failure for Bayesian estimators. The third chapter considers how identification of nonlinear solutions compares with that of linear solutions.

Chapter 1

The Statistical Implications of Common Identifying Restrictions for DSGE Models

Abstract. I reveal identification failure in a well-known dynamic stochastic general equilibrium (DSGE) model, and study the statistical implications of common identifying restrictions in this context. First, I provide a fully analytical methodology for determining all observationally equivalent values of the structural parameters in any parameter space. I show that parameter admissibility or sign restrictions may yield global identification for some parameter realizations, but not for others. Second, I derive a "plug-in" maximum likelihood estimator, which requires no numerical search. I use this tool to demonstrate that the specific identifying restriction impinges on both the location and distribution of the small-sample MLE in an idiosyncratic manner. Using this fact, I show how to compute correctly sized confidence intervals, and generalize the methodology for application to medium-scale models, such as Smets and Wouters (2007).

1   Introduction

DSGE models are the workhorse of modern macroeconomics. They are taught in nearly all graduate economics programs, and are a core empirical tool of

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

In this paper, I investigate whether it is possible to distinguish between parameter values in DSGE models on the basis of the data, what may be done when this is not case, and finally, what the statistical implications of the corrective actions are. I conduct this analysis using the very well-known model utilized in An and Schorfheide (2007). Although relatively small in scale, this model includes features of the seminal frameworks of Smets and Wouters (2003), Woodford (2003), and Christiano et al. (2005). Furthermore, these baseline specifications are the starting point for recent extensions, including the introduction of nonlinearity, non-normality, and more richly developed labor and financial markets.[5] In the penultimate section of the paper, I discuss how to generalize the methodology to larger-scale models, such as Smets and Wouters (2007).

The main findings are three-fold. First, I confirm that the conditional iden-

---

[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 May 2014, the representative paper of Smets and Wouters (2003) has 2,798 scholarly citations on Google Scholar.

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

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

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

[5]Galí et al. (2011) include unemployment, Bianchi (2013) allows for regime-switching, and Doh (2011), van Binsbergen et al. (2012), and Rudebusch and Swanson (2012) consider nonlinearity and the term structure. I discuss the implications of nonlinearity for identification in Chapter 3.

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

An obligation of any science is not only to produce groundbreaking results, but also to ensure that current knowledge is sound. Beyer and Farmer (2006) and Canova and Sala (2009) were two of the first to point out that the DSGE models regularly being estimated were not identified. Since the realization of this problem, analysts have been careful to restrict their analysis such that identification is ensured. Yet, the affect of these restrictions on confidence intervals has not been considered; a Classical estimator reported without valid confidence intervals is meaningless. Addressing this next phase of inquiry is the purpose of this paper.

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

2 Definitions

DSGE models are simply cross-equation and exclusion restrictions on systems of time series. Let Y be a $T n_Y \times 1$ vector of T observations of the $n_Y \times 1$ vector

---

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

[7] For example, one salient admissibility constraint is that the discount factor $\beta < 1$. A sign restriction is that a decrease in interest rates causes output to increase.

[8] The implications of identification failure for Bayesian estimators is discussed in Chapter 2.

of data Y_t and θ be an $n_\theta \times 1$ vector structural parameter. For each $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, define a continuous likelihood function $\ell(\theta; Y)$. Rothenberg (1971) gives the following definitions:[9]

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

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

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

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

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

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

---

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

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

Reduced form parameters are useful when they are available, since they completely characterize the likelihood; $\ell(\theta; Y) = \ell^*(\Pi; Y)$. For example, if Π is a reduced form parameter, and if $g(\theta_0) = g(\theta_0^*)$, then $\ell(\theta_0; Y) = \ell^*(g(\theta_0); Y) = \ell^*(g(\theta_0^*); Y) = \ell(\theta_0^*; Y)$ for all $Y \in \mathbb{R}^{Tn_y \times 1}$ by identity. In fact, using this observation, we have the following immediate corollary to Definitions 1 and 4.

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

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

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

An important preliminary point of clarification is why global identification of DSGE models matters in the first place. One reason is that observational equivalence can cause there to be multiple likelihood-maximizing parameter values in the admissible parameter space for a given sample. Say that $n_\theta = 1$, and the likelihood maximizer in a specific sample $Y \in \mathbb{R}^{Tn_Y \times 1}$ is $\theta_0 = \arg\max \ell(\theta; Y)$

$\ell(\theta; Y)$

θ_0 θ_1 $\mathbb{R}^1$
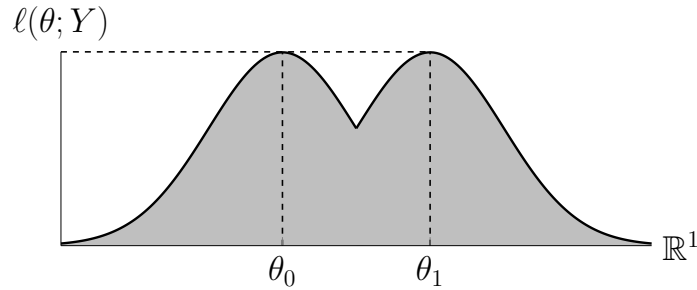
Figure 1.1: Likelihood for data sample $Y \in \mathbb{R}^{Tn_Y \times 1}$ under observational equivalence between likelihood maximizers $\theta_0 = \arg\max \ell(\theta; Y)$ and $\theta_0^* = \arg\max \ell(\theta; Y)$, but $\theta_0^* \neq \theta_0$.

where $\theta \in \mathbb{R}^1$. However, this value θ_0 has one observationally equivalent point, $\theta_0^* \neq \theta_0$ but $\theta_0^* \in \mathbb{R}_1$. Thus, $\theta_0^* = \arg\max \ell(\theta; Y)$ by definition, and the likelihood will have two equally tall peaks in the domain $\mathbb{R}^1$, as in Figure 1.1. This would cause the results of numerical search for a unique maximum to be misleading. Yet it is important to take note that conversely, two peaks of the likelihood for a given Y necessarily implies neither observational equivalence nor lack of global identification. Rather, this equality is only necessarily a property of the specific sample Y. In another sample, that property might change.

Next, I utilize these definitions to place my paper in context of the wider literature.

3   Contributions to the Literature

This paper is part of a small, but important literature which considers identification of DSGE models. Following the realization of problems by Beyer and Farmer (2006) and Canova and Sala (2009), conditions for local identification were provided by Iskrev (2010), Komunjer and Ng (2011), and Qu and Tkachenko (2012). However, local identification is merely necessary, and not sufficient for global identification; in terms of Definitions 2 and 3, there may be observationally equivalent structural parameters outside of an open neighborhood of the point of interest. In general, it is very difficult to state necessary and sufficient conditions for global identification in nonlinear models, and only the overly strong sufficient

conditions of Gale and Nikaidô (1965) are typically useful. As a result, there exist no published results on global identification of DSGE models, although working papers by Fukač et al. (2007), Kocięcki and Kolasa (2013), and Qu and Tkachenko (2013) have managed to make strides in this direction.

The three aforementioned papers on global identification are the most generally useful contributions on the topic thus far. Yet, there are a number of significant caveats yet to be overcome. All existing approaches suggest some form of searching a given parameter space for observationally equivalent points. Such a numerical search is daunting even in a small multidimensional parameter space, and infeasible in a desirably large one. Furthermore, the computational toll necessary to search for even one observationally equivalent point makes repeating the process for many points impractical. In addition, the numerical algorithms necessary to solve DSGE models insert small errors in the mapping from structural parameters to moments which may lead two observationally equivalent points to seem distinct, or vice-versa.

The ultimate reason for these difficulties is that DSGE models neither have analytical solutions nor reduced form representation in general. The complications are a preliminary assumption of all of the papers of Komunjer and Ng (2011), Kocięcki and Kolasa (2013), and Qu and Tkachenko (2013), all of which use the model of An and Schorfheide (2007) to demonstrate their methodologies. In this paper, I do not offer generally applicable results for assessing global identification. Rather, I simply point out that main model that has been studied in this literature, An and Schorfheide's, has both an analytical solution *and* reduced form representation. These two facts allow me to address all of the problems discussed above using simple econometric tools, and the resulting discussion is both concise and transparent. Finally, this observation will allow me to demonstrate how identifying restrictions affect the distribution of the MLE, the main concern of this paper.

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

4   Solution and Reduced Form Representation

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

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \phi_{rr} & \phi_{ry} & \phi_{r\pi} \\ \phi_{yr} & \phi_{yy} & \phi_{y\pi} \\ \phi_{\pi r} & \phi_{\pi y} & \phi_{\pi\pi} \end{bmatrix}}_{\Phi} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \quad \Omega \equiv E(U_t U_t') = \begin{bmatrix} \omega_r^2 & \cdot & \cdot \\ \omega_{yr} & \omega_y^2 & \cdot \\ \omega_{\pi r} & \omega_{\pi y} & \omega_\pi^2 \end{bmatrix} \quad (1.1)$$

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

---

[11] D_{n_Y} is the duplication matrix, $D_{n_Y}^+ = (D_{n_Y}' D_{n_Y})^{-1}D_{n_Y}'$ is its Moore-Penrose pseudoinverse, and $\otimes$ is the Kronecker product. See Abadir and Magnus (2005).

Table 1.1: Model and parameter names.

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

addition, efficient estimators $\widehat{\Phi}$ and $\widehat{\Omega}$ are trivially available from ordinary least squares.

The unrestricted Gaussian VAR(1) with 15×1 reduced form parameter $\Pi = (\text{vec}(\Phi)', \text{vech}(\Omega)')'$ has many desirable properties. But how does it compare with a standard DSGE model? The linearized equilibrium equations for the An and Schorfheide (2007) model are given by the following 6 equilibrium equations. A complete derivation is available in Appendix B. Variable and parameter names are given in Table 1.1.

$$z_t = \rho_z z_{t-1} + \varepsilon_{zt} \tag{1.2}$$

$$g_t = \rho_g g_{t-1} + \varepsilon_{gt} \tag{1.3}$$

$$r_t = \rho_r r_{t-1} + (1 - \rho_r)\psi_\pi \pi_t + (1 - \rho_r)\psi_y(y_t - g_t) + \varepsilon_{rt} \tag{1.4}$$

$$y_t = E_t y_{t+1} + g_t - E_t g_{t+1} - (1/\tau)(r_t - E_t \pi_{t+1} - E_t z_{t+1}) \tag{1.5}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - g_t) \tag{1.6}$$

$$c_t = y_t - g_t \tag{1.7}$$

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

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

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

$$\Sigma_\varepsilon(\theta) \equiv E(\varepsilon_t \varepsilon_t'; \theta) = \begin{bmatrix} \sigma_z^2 & \cdot & \cdot \\ \sigma_{gz} & \sigma_g^2 & \cdot \\ \sigma_{rz} & \sigma_{rg} & \sigma_r^2 \end{bmatrix} \tag{1.8}$$

In this paper I allow for the possibility that the off-diagonal covariances are nonzero; recall, these are related to correlations, for example, by $\rho_{gz} = \sigma_{gz}/(\sigma_g \sigma_z)$. θ is the $n_\theta = 16$ dimensional column vector structural parameter

$$\underset{(16 \times 1)}{\theta} = (\tau, \beta, \nu, \phi, \Pi, \psi_\pi, \psi_y, \rho_z, \rho_g, \rho_r, \sigma_z, \sigma_g, \sigma_r, \sigma_{gz}, \sigma_{rz}, \sigma_{rg})' \tag{1.9}$$

Respectively, Equations (1.2)-(1.7) are rules of motion for TFP and government spending, a Taylor rule, Euler equation, Phillips curve, and aggregate accounting equality. At first blush, the resemblance between this model and the VAR in Equation (1.1) is not obvious. In particular, very specific assumptions are necessary for a given DSGE model to have reduced form representation (See Ravenna (2007) and Giacomini (2013)). Furthermore, in order to form any reduced form representation, it is usually necessary to use numerical solution algorithms like Sims (2002)'s. Such a black box makes it impossible to exploit the certain identifiability of the VAR parameters for the purposes of determining the identification of the structural parameters.

In the next section, I show how a simplified version of the An and Schorfheide model may be solved analytically, and that this solution is in fact a special case of Equation (1.1). Then, I show the same for the full model.

4.1 Simplified An and Schorfheide Model

Consider the special case of Equations (1.2)-(1.7) in which TFP is iid, $z_t = \varepsilon_{zt}$, and government spending is zero, $g_t = 0 \; \forall \; t$. This model has three equations,

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

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

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

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

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

Redefine $\rho_r = \phi_{rr}$ and $\varepsilon_{rt} = d_{rz}\varepsilon_{zt} + d_{rg}\varepsilon_{gt} + d_{rr}\varepsilon_{rt}$. In addition, let inflation be augmented by a shock, $\pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \varepsilon_{\pi t}$, and define $\varepsilon_{yt} = (1/\tau)\varepsilon_{zt}$. Thus, the above three equations become the following:[12]

$$r_t = \rho_r r_{t-1} + \varepsilon_{rt} \tag{1.10}$$

$$y_t = E_t y_{t+1} - (1/\tau)(r_t - E_t \pi_{t+1}) + \varepsilon_{yt} \tag{1.11}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \varepsilon_{\pi t} \tag{1.12}$$

The three exogenous variables ε_{rt}, ε_{yt}, and $\varepsilon_{\pi t}$ are iid mean-zero Gaussian innovations. While ε_{rt} and ε_{yt} are respectively the idiosyncratic portion of monetary policy and a technological innovation, $\varepsilon_{\pi t}$ may be conceptualized as a cost-push shock or observational error, for instance (See Ireland (2004)). Allowing for the possibility that three innovations are mutually correlated, the 3×1 vector of in-

---

[12]In other words, the reduced form rule of motion for interest rates, Equation (1.10) is equivalently the minimal state variable solution of a Taylor rule that responds to both inflation and output when r_{t-1} is the only lagged endogenous variable in the model; see also McCallum (1983) and McCallum (1999). Thus, Equation (1.10) is simply a reparameterization of the Taylor rule and this model will not produce sunspots.

novations $\varepsilon_t = [\varepsilon_{rt}, \varepsilon_{yt}, \varepsilon_{\pi t}]'$ has variance-covariance matrix

$$\Sigma_\varepsilon(\theta) \equiv E(\varepsilon_t \varepsilon_t'; \theta) = \begin{bmatrix} \sigma_r^2 & \cdot & \cdot \\ \sigma_{yr} & \sigma_y^2 & \cdot \\ \sigma_{\pi r} & \sigma_{\pi y} & \sigma_\pi^2 \end{bmatrix} \tag{1.13}$$

where θ is the 10×1 vector structural parameter

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

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

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \rho_r & 0 & 0 \\ \phi_{yr} & 0 & 0 \\ \phi_{\pi r} & 0 & 0 \end{bmatrix}}_{\Phi(\theta)} \underbrace{\begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix}}_{} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \tag{1.14}$$

$$\underbrace{\begin{bmatrix} \omega_r^2 & \cdot & \cdot \\ \omega_{yr} & \omega_y^2 & \cdot \\ \omega_{\pi r} & \omega_{\pi y} & \omega_\pi^2 \end{bmatrix}}_{\Omega(\theta)} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ \phi_{yr}/\rho_r & 1 & 0 \\ \phi_{\pi r}/\rho_r & \kappa & 1 \end{bmatrix}}_{D(\theta)} \underbrace{\begin{bmatrix} \sigma_r^2 & \cdot & \cdot \\ \sigma_{yr} & \sigma_y^2 & \cdot \\ \sigma_{\pi r} & \sigma_{\pi y} & \sigma_\pi^2 \end{bmatrix}}_{\Sigma_\varepsilon(\theta)} \underbrace{\begin{bmatrix} 1 & \phi_{yr}/\rho_r & \phi_{\pi r}/\rho_r \\ 0 & 1 & \kappa \\ 0 & 0 & 1 \end{bmatrix}}_{D(\theta)'} \tag{1.15}$$

where

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

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

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

Table 1.2: Simplified An and Schorfheide model candidate calibration θ_0 and theoretically motivated parameter space Θ_0. $\varepsilon =$1e-6. Note, parameter bounds are purposefully allowed to be generous; see also Table 1.1 for microfounded definitions.

	Param	Lower	θ_0	Upper
1	τ	0.1	2	3.5
2	β	0.975	0.9975	1-ε
3	κ	ε	0.33	3
4	ρ_r	ε	0.75	1-ε
5	σ_r	ε	2e-2	1
6	σ_y	ε	2e-2	1
7	σ_π	ε	2e-2	1
8	σ_{yr}	-1	1e-4	1
9	$\sigma_{\pi r}$	-1	1e-4	1
10	$\sigma_{\pi y}$	-1	-1e-4	1

parameter

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

It will ultimately prove useful to discuss specific values of the structural parameters. Table 1.2 provides a candidate calibration θ_0, along with upper and lower bounds for the individual elements of θ; these make up the theoretically justifiable population parameter space Θ_0. One possible interpretation of Θ_0 is as the boundaries of diffuse priors over θ. Another is simply the support of any bounded prior, or the maximum admissible space. At θ_0,

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

4.2 Full An and Schorfheide Model

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

symbolic computation.

Recall that the structural parameter for the full model is the 16×1 vector given in Equation (1.9). While the solution of the simplified model followed from the fact that only lagged interest rates appeared in the equilibrium conditions, the same is not true here. To approach the solution of the full model analytically, note that the minimal solution of the model has the following general form (See Komunjer and Ng (2011) and Kailath et al. (2000)):

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

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

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

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

Plugging in the corresponding equations for $E_t y_{t+1}$ and $E_t \pi_{t+1}$, along with $E_t z_{t+1} = \rho_z z_t$, into aggregate demand, Equation (1.5),

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

Plugging the corresponding equation for $E_t \pi_{t+1}$, along with the expression for y_t

just derived into the Phillips curve, Equation (1.6),

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

$$+ \underbrace{\left(\kappa c_{yr} + \left(\beta + \frac{\kappa}{\tau}\right)c_{\pi r} - \frac{\kappa}{\tau}\right)}_{f_{\pi r}} r_t$$

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

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ f_{yz} & f_{yg} & f_{yr} \\ f_{\pi z} & f_{\pi g} & f_{\pi r} \end{bmatrix}}_{F(\theta)} \underbrace{\begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix}}_{X_t} \tag{1.16}$$

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

$$\underbrace{\begin{bmatrix} c_{rz} & c_{rg} & c_{rr} \\ c_{yz} & c_{yg} & c_{yr} \\ c_{\pi z} & c_{\pi g} & c_{\pi r} \end{bmatrix}}_{C(\theta)} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ f_{yz} & f_{yg} & f_{yr} \\ f_{\pi z} & f_{\pi g} & f_{\pi r} \end{bmatrix}}_{F(\theta)} \times \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{rz} & c_{rg} & c_{rr} \end{bmatrix}}_{A(\theta)} \tag{1.17}$$

Since the elements of F have been expressed in terms of the elements of C and θ, the system in Equation (1.17) yields 9 equations and 9 unknowns, the elements of C. Although it is infeasible to solve for these as functions of θ by hand, it is straightforward to make use of symbolic computation software for this purpose. Solving the model using MATLAB's built-in symbolic computation software reveals that there are exactly three solutions. However, only one of these solutions implies a stable solution at θ_0, as judged by the modulus of the eigenvalues of $A(\theta_0)$. The generalized functional form of this unique stable solution is

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

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

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

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

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

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

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

As in the simplified model, θ is assigned a candidate calibration θ_0, and parameter space Θ_0, in Table 1.3. It is verified that the values of this analytical ABCD solution correspond to Sims (2002)'s numerical solution at the same point. However, the state space parameters are not reduced form parameters. Thus, it is desirable to explore any companion forms the model might have. It is now useful to state a

Table 1.3: Full An and Schorfheide candidate calibration θ_0 and space Θ_0. ε=1e-6.

	Param	Lower	θ_0	Upper
1	τ	0.1	2	3.5
2	β	0.975	0.9975	1-ε
3	ν	ε	0.1	1
4	ϕ	50	53.68	60
5	Π	1+ε	1.008	1.03
6	ψ_π	-1	1.5	3
7	ψ_y	-1	0.125	1.25
8	ρ_z	ε	0.9	1-ε
9	ρ_g	ε	0.95	1-ε
10	ρ_r	ε	0.75	1-ε
11	σ_z	ε	3e-2	1
12	σ_g	ε	6e-2	1
13	σ_r	ε	2e-2	1
14	σ_{gz}	-1	1e-4	1
15	σ_{rz}	-1	1e-4	1
16	σ_{rg}	-1	-1e-4	1

simple result.

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

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

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

Without any exogenous restrictions, the An and Schorfheide model satisfies $Y_t = F(\theta)X_t$ from Equation (1.16). Thus, its observables Y_t have restricted VAR(1) representation. Defining U_t to be a 3×1 vector of nonstructural innovations $D\varepsilon_t$ with covariance matrix $E(U_tU_t') \equiv \Omega(\theta)$, by the above result, the observables Y_t have the following VAR(1) reduced form representation, another special case of

Equation (1.1):

$$
\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \phi_{rr} & 0 & \phi_{r\pi} \\ \phi_{yr} & \rho_g & \phi_{y\pi} \\ \phi_{\pi r} & 0 & \phi_{\pi\pi} \end{bmatrix}}_{\Phi(\theta)} \underbrace{\begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix}}_{} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t}
\tag{1.18}
$$

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

$$
\underbrace{\begin{bmatrix} \phi_{rr} & 0 & \phi_{r\pi} \\ \phi_{yr} & \rho_g & \phi_{y\pi} \\ \phi_{\pi r} & 0 & \phi_{\pi\pi} \end{bmatrix}}_{\Phi(\theta)} = \underbrace{\begin{bmatrix} c_{rz} & 0 & c_{rr} \\ c_{yz} & \rho_g & c_{yr} \\ c_{\pi z} & 0 & c_{\pi r} \end{bmatrix}}_{C(\theta)} \times \underbrace{\begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ c_{rz} & 0 & c_{rr} \end{bmatrix}}_{A(\theta)} \times \underbrace{\begin{bmatrix} c_{rz} & 0 & c_{rr} \\ c_{yz} & \rho_g & c_{yr} \\ c_{\pi z} & 0 & c_{\pi r} \end{bmatrix}^{-1}}_{C(\theta)^{-1}}
\tag{1.19}
$$

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

$$\tag{1.20}$$

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

$$
\underset{(13\times1)}{\Pi(\theta)} = (\phi_{rr}, \phi_{yr}, \phi_{\pi r}, \rho_g, \phi_{r\pi}, \phi_{y\pi}, \phi_{\pi\pi}, \omega_r, \omega_y, \omega_\pi, \omega_{yr}, \omega_{\pi r}, \omega_{\pi y})'
$$

At θ_0, Equation (1.18) has the following realization:

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

5 Identification Analysis

As motivated previously, the global identification of the reduced form VAR(1) Equation (1.1) is trivial, and furthermore, both the simplified and full versions of the model are now known to have restricted VAR(1) representation. Yet, the iden-

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

Figure 1.2: Simplified model functional mapping $g : \theta \to \Pi$.

tification of θ is the relevant issue. In this section I study the identifiability of the structural parameters θ from the reduced form parameters Π.

5.1 Simplified Model

In order to distinguish whether θ is globally identified in the real plane, it is sufficient to show that θ is uniquely recoverable from Π regardless of θ's realization in the reals. Although the mapping from structural parameters θ to reduced form parameters Π is nonlinear, by virtue of the analytical solution developed in the preceding section, the correspondence is closed-form. In place of the complicated functions themselves, $\times$'s are used in Figure 1.2 to represent the dependence of each element of Π on each element of θ. For instance, the standard error ω_π is a function of all 10 structural parameters. As noted in the figure, it will henceforth clarify the analysis to name this functional dependence $\Pi = g(\theta)$. A similar identification table was also recently utilized by Hamilton and Wu (2012) as a preliminary step in investigating the identification of affine term structure models.

The number of structural parameters, $n_\theta = 10$ columns, exceeds the number of reduced form parameters, $n_\Pi = 9$ rows; in other words, it is immediately apparent that the necessary order condition for identification, $n_\Pi \geq n_\theta$, is violated. Thus, at least one of the structural parameters must be set to a constant for the complement subset to be conditionally identified. In order to distinguish which

structural parameter is such a candidate, or if more than one parameter must be set, one possible procedure is to check local identification. As motivated by Iskrev (2010), a reasonable starting point is to compute the Jacobian $J(\theta_0) = \partial\Pi/\partial\theta'|_{\theta=\theta_0}$ and successively eliminate columns until those that remain are full column rank; the dropped columns correspond to parameters to be set. However, such an approach is only valid at θ_0 and overlooks the issue of global identification more broadly.

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

So, ϑ is globally identifiable at any given point in any parameter space if $n_\vartheta = n_\Pi$ and g is injective. The latter requirement is verified if g^{-1} exists and is unique. Thus, a central question is whether a unique inverse of the specific functional form of g exists.

By examining the functional form of the vector-valued function $g(\theta)$ represented by Table 1.2, it appears that the parameter σ_π is a reasonable candidate for the role of $\overline{\alpha}$. This implies that the parameter to be analyzed using the analytical approach is

$$\underset{(9\times1)}{\vartheta} = (\tau, \beta, \kappa, \rho_r, \sigma_r, \sigma_y, \sigma_\pi, \sigma_{yr}, \sigma_{\pi r}, \sigma_{\pi y})'$$

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

Table 1.4: Simplified model observational equivalence in $\mathbb{R}^9$ but not Θ_0. $\Pi = g(x; \overline{\sigma}_\pi)$; $x = \vartheta_0$ or ϑ_0^*.

		$g(x; \overline{\sigma}_\pi)$		ϑ_0	ϑ_0^*
1	ρ_r	0.75	τ	2	2
2	ϕ_{yr}	1.55	β	0.9975	**1.51**
3	$\phi_{\pi r}$	2.04	κ	0.33	**-0.17**
4	ω_r	2e-2	ρ_r	0.75	0.75
5	ω_y	5e-2	σ_r	2e-2	2e-2
6	ω_π	6e-2	σ_y	2e-2	2e-2
7	ω_{yr}	9e-4	σ_{yr}	1e-4	1e-4
8	$\omega_{\pi r}$	12e-4	$\boldsymbol{\sigma_{\pi r}}$	1e-4	**2e-4**
9	$\omega_{\pi y}$	28e-4	$\boldsymbol{\sigma_{\pi y}}$	-1e-4	**1e-4**

expression for τ is

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

and ρ_r is itself one of the reduced form parameters. The remaining expressions for σ_r, $\sigma_{\pi r}$, and σ_{yr} are closed-form, but too complicated to provide any intuition. The most important result, however, has to do with the remaining 4 elements of ϑ. Even after normalizing the relationship between standard errors and variances, for any Π, there are exactly 2 expressions for each of β, κ, $\sigma_{\pi r}$, and $\sigma_{\pi y}$ in terms of Π, no more, no less. In other words, g is *not* injective, and its inverse yields two values:

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

There are several important consequences of Table 1.4. First, ϑ_0 and ϑ_0^* are observationally equivalent despite that the 9×9 Jacobian $J(\vartheta_0; \overline{\sigma}_\pi) = \partial\Pi(\vartheta; \overline{\sigma}_\pi)/\partial\vartheta'|_{\vartheta=\vartheta_0}$ is full rank, a necessary and sufficient condition for local identification. Second, the value ϑ_0^* – and specifically, the values of β and κ at ϑ_0^* – fall outside of Θ_0. Therefore, an assessment of global identification at ϑ_0 based on numerically searching over Θ_0 would conclude simply that ϑ is globally identified in Θ_0, whereas this conclusion is possibly misleading of the fact that there is an observationally equivalent ϑ_0^* in $\mathbb{R}^9$ outside of, but relatively close to, Θ_0. Third,
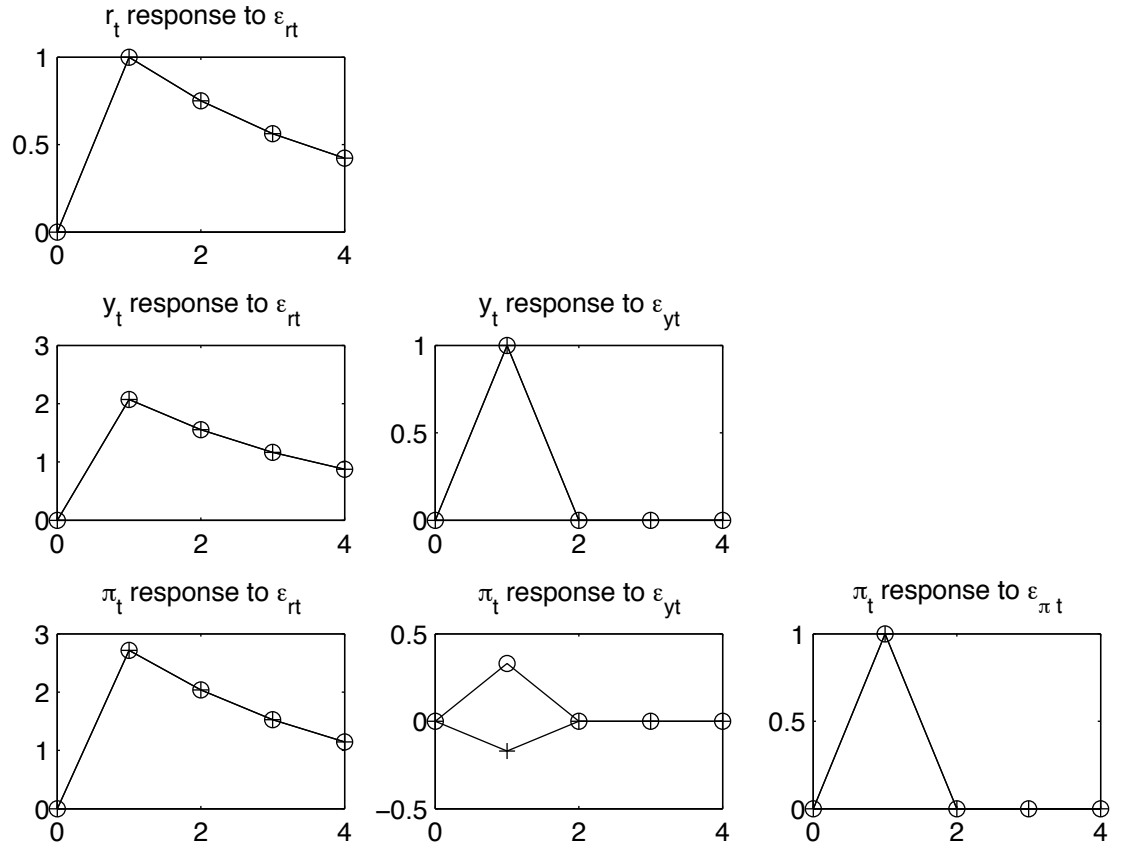
Figure 1.3: Simplified model impulse-responses for observationally equivalent points: ϑ_0 ($\circ$) vs ϑ_0^* ($+$). Triangularity of impulse-responses reflects triangularity of $D(\theta)$ in Equation (1.15).

computational time is a fraction of a second.

A natural next question is whether the economic implications of both ϑ_0 and ϑ_0^* are the same. I provide impulse-responses for both points in Figure 1.3. When the crosshairs corresponding to ϑ_0^* exactly hit the bullseyes corresponding to ϑ_0, the economic implications are equivalent. In only one case, π_t's response to ε_{yt}, do the economic implications of each point differ. In fact, this can be traced back to Equation (1.15) as arising from the difference in κ between points. There are two important implications of this observation. First, the monetary policy impulse-responses in the first column of Figure 1.3 are robust to observational equivalence. This means that if those impulse responses were the only object of economic interest, a valid normalization is to simply drop ϑ_0^*. Second, the economic implications of the two points for π_t's response to ε_{yt} differ. Therefore, if this is the object of economic analysis, simply dropping ϑ_0^* is not benign.

The good news is that is reasonable to distinguish between ϑ_0 and ϑ_0^* on the basis of both the bounded priors embodied in Θ_0, and sign restrictions. For instance, not only is $\kappa < 0$ contrary to standard theory, it also implies inflation will decrease following a cost-push shock (to ε_{yt}). In addition, $\beta > 1$ is certainly disconcerting. Specifically, if the space in which global identification is important is Θ_0, then Table 1.4 has simply shown us that ϑ_0 is globally identified in Θ_0. Thus, although this model does engender observational equivalencies, the two facts that global identification is defined in terms of a parameter space, and that macroeconomists have a good idea of what parameter spaces are important, allow the analyst to eliminate the nuisance ϑ_0^* in this case.

Yet, it is not always possible to distinguish between observationally equivalent points on the basis of the bounds of Θ_0 alone. Since evaluating global identification for a point may be done very efficiently, I am able to search the parameter space Θ_0 for values ϑ_1 which have an observationally equivalent ϑ_1^* that is also in Θ_0. Such an example is given in Table 1.5. The impulse responses for each point are given in Figure 1.4. Again, only π_t's response to ε_{yt} differs from point to point. However, in this case, both impulse responses have the same sign, owing to the fact that the value for κ at each point is the same. Evidently, observational

Table 1.5: Simplified model observational equivalence in Θ_0. $\Pi = g(x; \overline{\sigma}_\pi)$; $x = \vartheta_0$ or ϑ_0^*.

		$g(x; \overline{\sigma}_\pi)$		ϑ_1	ϑ_1^*
1	ρ_r	0.93	τ	1.01	1.01
2	ϕ_{yr}	0.03	β	0.9836	**0.9980**
3	$\phi_{\pi r}$	1.002	κ	2.70	**2.25**
4	ω_r	0.41	ρ_r	0.93	0.93
5	ω_y	0.08	σ_r	0.41	0.41
6	ω_π	0.60	σ_y	0.07	0.07
7	ω_{yr}	0.03	σ_{yr}	0.03	0.03
8	$\omega_{\pi r}$	0.25	$\boldsymbol{\sigma_{\pi r}}$	-6.1e-3	**5.1e3**
9	$\omega_{\pi y}$	0.05	$\boldsymbol{\sigma_{\pi y}}$	-9e-4	**9e-4**

Table 1.6: Identification of ϑ in simplified model: 100,000 uniformly drawn points from Θ_0.

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

equivalence is a more trying issue at the realization ϑ_1, since neither it nor ϑ_1^* may be easily eliminated. More elaborate identifying restrictions are necessary.

Consider, then, the analyst who decides to simply choose ϑ_1^*, since the value of β is closer to the original value in θ_0 of 0.9975. Macroeconomists usually have very strong convictions about the value of this parameter in particular. This action would reflect, for example, the will of an analyst with a very tight prior for β centered at 0.9975. Depending on the analyst's ultimate inferential goal, many possible identifying restrictions are possible.

What is the likeliness of drawing a point like ϑ_0 versus a point like ϑ_1 in the parameter space Θ_0? Obviously, if there are relatively more ϑ_0, this is all else equal a good thing, since the analyst will be required to enforce relatively less stringent identifying restrictions. Table 1.6 presents the results of testing for local identification, and global identification in both Θ_0 and all $\Theta \subset \mathbb{R}^9$, at 100,000 uniformly drawn points from Θ_0. The only restriction I make on the points drawn is that the variance-covariance matrix $\Sigma_\varepsilon(\vartheta; \overline{\alpha})$ must be positive semidefinite. Recall, since $\Omega = D\Sigma_\varepsilon D'$ in Equation (1.15), Σ_ε positive semidefinite implies that so is Ω.
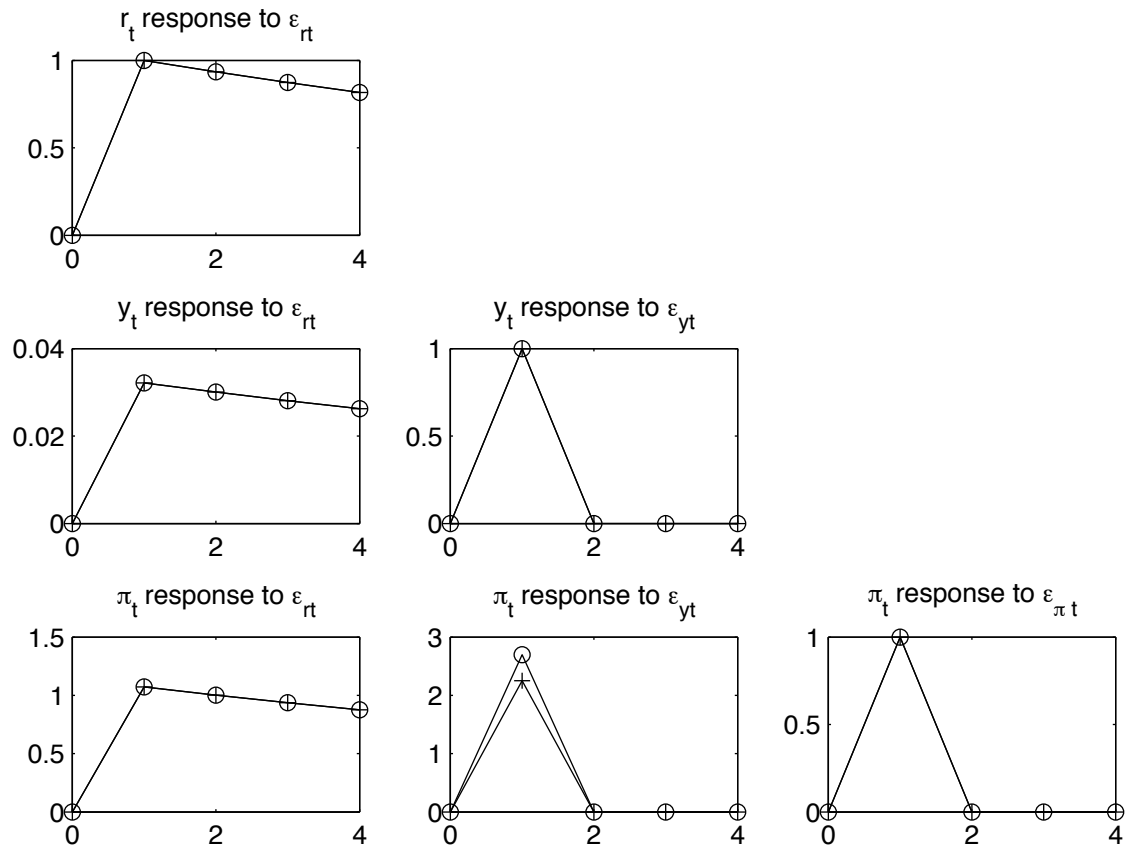
Figure 1.4: Simplified model impulse-responses for two theoretically plausible and observationally equivalent points: ϑ_1 ($\circ$) vs ϑ_1^* ($+$).

ϑ is locally identified at all points, but has a real-valued observationally equivalent point at all but 0.04% of draws.[13] Furthermore, the identifying restriction that ϑ must be in Θ_0 does not successfully yield a unique point in roughly a third of the space. Thus, stronger identifying restrictions like the previously discussed β criterion are frequently needed.

5.2 Plug-In Maximum Likelihood

While we are now aware of what types of identifying restrictions are necessary in the simple model, and when they must be implemented, we are not fully cognizant of the effects of these normalizations on confidence intervals. One approach to understand the statistical outcomes of a given normalization is to implement these identifying restrictions in the Monte Carlo distribution of a given estimator, such as the maximum likelihood estimator.[14] However, under normal circumstances, it would be impossible to construct such a distribution for DSGE models in particular. The reasons are three-fold. Firstly, without the knowledge of how many maxima are in $\mathbb{R}^9$, it would be impossible to determine how many maximum likelihood estimators exist for each draw i. Secondly, as demonstrated by Andreasen (2010), even the most sophisticated global algorithms for likelihood maximization, including simulated annealing and genetic numerical search, are prone to failure. Furthermore, search algorithms are only as accurate as the termination tolerance chosen by the analyst, which is particularly worrying given the topological characterization of weak identification is flatness in the likelihood surface (See Canova and Sala (2009)). Finally, likelihood maximization is computationally costly, particularly for the most reliable genetic algorithm, making it infeasible to compute the distribution for large N.

In fact, the mapping g^{-1} has other uses besides assessing identification; it may also be used as the basis of a "plug-in" maximum likelihood estimator. This tool is uniquely suited for computing the Monte Carlo distribution, and investi-

---

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

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

gating the statistical consequences of implemented normalizations. I motivate this estimator using the minimum chi-squared estimator, or MCSE, implemented by Hamilton and Wu (2012) and first suggested by Rothenberg (1973). When $\widehat{\Pi}$ is the maximum likelihood estimator of the reduced form parameters, and $\widehat{\mathcal{I}}$ is an efficient estimator of the information matrix with respect to Π, the MCSE is defined by the following:

$$\widehat{\vartheta}_{MCSE} = \arg\min_{\vartheta \in \Theta} \ (\widehat{\Pi} - g(\vartheta; \overline{\alpha}))'\widehat{\mathcal{I}}(\widehat{\Pi} - g(\vartheta; \overline{\alpha})) \tag{1.21}$$

As Hamilton and Wu (2012) go on to show, this estimator is asymptotically efficient. In addition, in the special case that $n_\vartheta = n_\Pi$, the minimal value of the criterion is zero, in which case the analyst may simply use an identity matrix as the weighting matrix, and find the MCSE as equivalently the minimizer of

$$(\widehat{\Pi} - g(\vartheta; \overline{\alpha}))'(\widehat{\Pi} - g(\vartheta; \overline{\alpha})) \tag{1.22}$$

Notice, since (1.22) a quadratic form, $\widehat{\vartheta}_{MLE}$ is equivalently defined by $\widehat{\Pi} = g(\widehat{\vartheta}_{MLE}; \overline{\alpha})$. In other words, given any data set, under the conditional identification scheme $\overline{\alpha} = \overline{\sigma}_\pi$ there are exactly two points which maximize the likelihood:

$$g^{-1}(\widehat{\Pi}; \overline{\sigma}_\pi) \Big\langle \begin{array}{c} \widehat{\vartheta} \\ \\ \widehat{\vartheta}^* \end{array}$$

The coefficients $\widehat{\Pi}$ are available by restricted feasible generalize least squares (FGLS).[15] In addition, given g^{-1} is known in closed-form, the analyst need only plug-in to obtain both points $\widehat{\vartheta}$ and $\widehat{\vartheta}^*$ that maximize the likelihood. As in Hamilton and Wu (2012), there is no uncertainty that each of $\widehat{\vartheta}$ and $\widehat{\vartheta}^*$ maximizes the likelihood. My extension of deriving g^{-1} directly also implies there is no uncertainty about the number of likelihood maxima or the reliability of search algorithms, nor

---

[15]In Hamilton and Wu (2012), ordinary least squares is utilized. In the current case, the exclusion restrictions on Φ require restricted FGLS. See Lutkepohl (2005) p. 20. GLS is equivalent to the MLE under normality of U_t. See Hamilton (1994) p. 222.
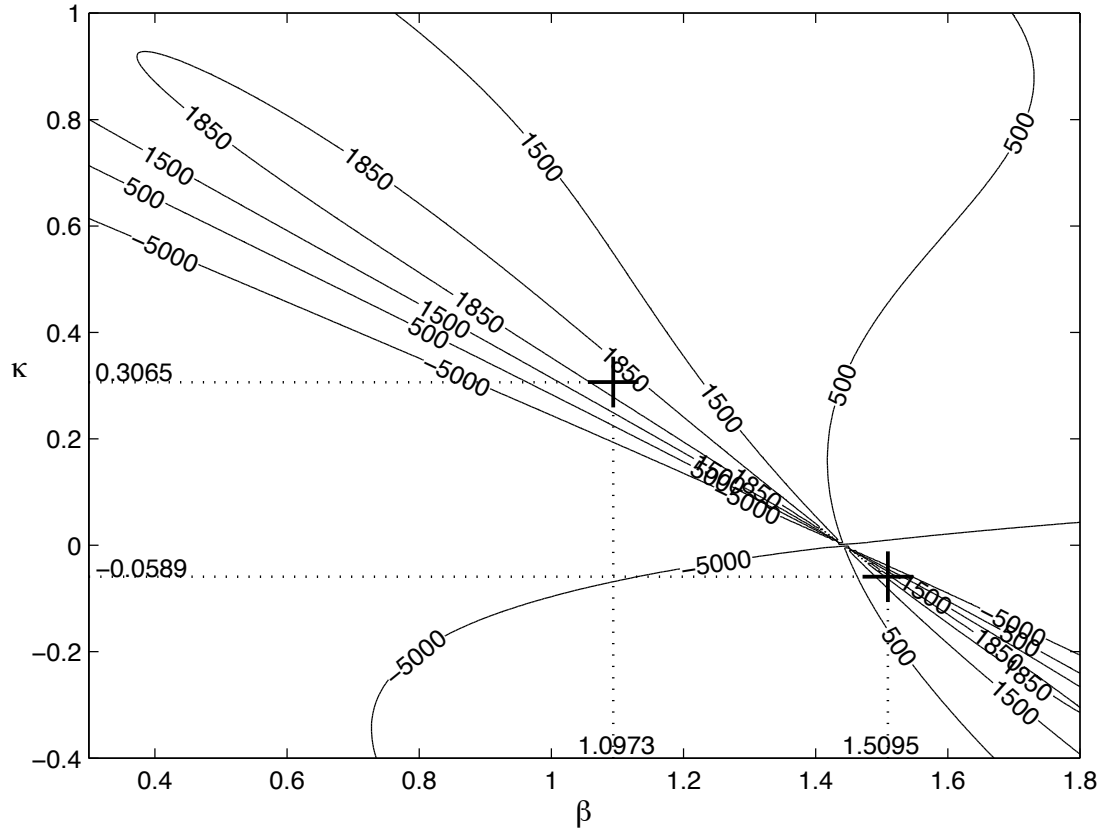
Figure 1.5: Contour plot of log-likelihood. Maximum achieved at identical 1909.9 for exactly two points +, which are the plug-in MLEs. Compare with depicted bimodal univariate likelihood in Figure 1.1.

is there computational toll from numerical search.

To verify, I use the value ϑ_0 to generate a data set of length $T = 250$ and reestimate the parameters of the model using the plug-in technique. As predicted previously in Table (1.4), for instance, this estimator yields two different values for β and κ. By computing the contour plot of the likelihood over these values, it shown that the plug-in MLEs do maximize the likelihood function, with exactly equal values at both points.

5.3 Small Sample Distribution of MLE

It is now feasible to compute a Monte Carlo simulation of the small-sample standard errors of the MLE.[16] First, I consider the model subject to no identifying restrictions is given in Figure 5.3 with $T = 250$ and N=10,00. This T is equivalent to 62.5 years of quarterly data, approximately the available window of post-war data on interest rates, output, and inflation. The entire distribution is computed in a few seconds. To understand how this distribution is computed requires first an understanding of the sample. While roughly 78% of the draws $\widehat{\Pi}_i$ yield two real-valued estimators $\widehat{\vartheta}_i$ and $\widehat{\vartheta}_i^*$, 22% yield two complex-valued estimators, and 0% yield only one real-valued point. There are two conflicting perspectives on how to deal with the complex-valued estimators. One is to include these in the Monte Carlo distribution. The other is to simply try to find a local maximum in the reals. However, whereas the first approach takes this analysis out of the comfort zone of meaningful macroeconomic analysis, the second suggests an estimator known not to maximize the likelihood. Since there is no formal econometric basis for distinguishing between these two options, for the 22% of draws yielding two complex points, I take the safest route possible and simply say the estimator does not exist. Thus, the restricted estimator is technically a restricted MLE. When there are two real-valued estimators, each receives a weight of one in the distribution. The pseudo-code for Figure 5.3 follows:

1. For draw $\widehat{\Pi}_i$, compute both estimators $\widehat{\vartheta}_i$ and $\widehat{\vartheta}_i^*$ by plugging in to g^{-1}.

2. If both are outside of reals, the estimator is said to not exist for the given draw (22% of utilized sample).

---

[16]Qu (2013), Guerron-Quintana et al. (2013), Dufour et al. (2013), and Andrews and Mikusheva (2013) have all discussed the issue of weak identification, and how to compute correctly sized large-sample confidence intervals and tests. Since the distributions I consider are bootstrapped small sample standard errors, they do not rely on the asymptotic approximations that break down under weak identification (See Stock et al. (2002), for example). Note, it is not generally possible to compute a bootstrapped distribution for weakly identified models. This is due to the fact that the bootstrap is known to be invalid under weak identification, an outcome related to the failure of Edgeworth expansion in this case (See Hall (1992)). However, the bootstrap statistics I calculate are applied only to the restricted FGLS estimator of the reduced form VAR parameters, which are strongly identified.
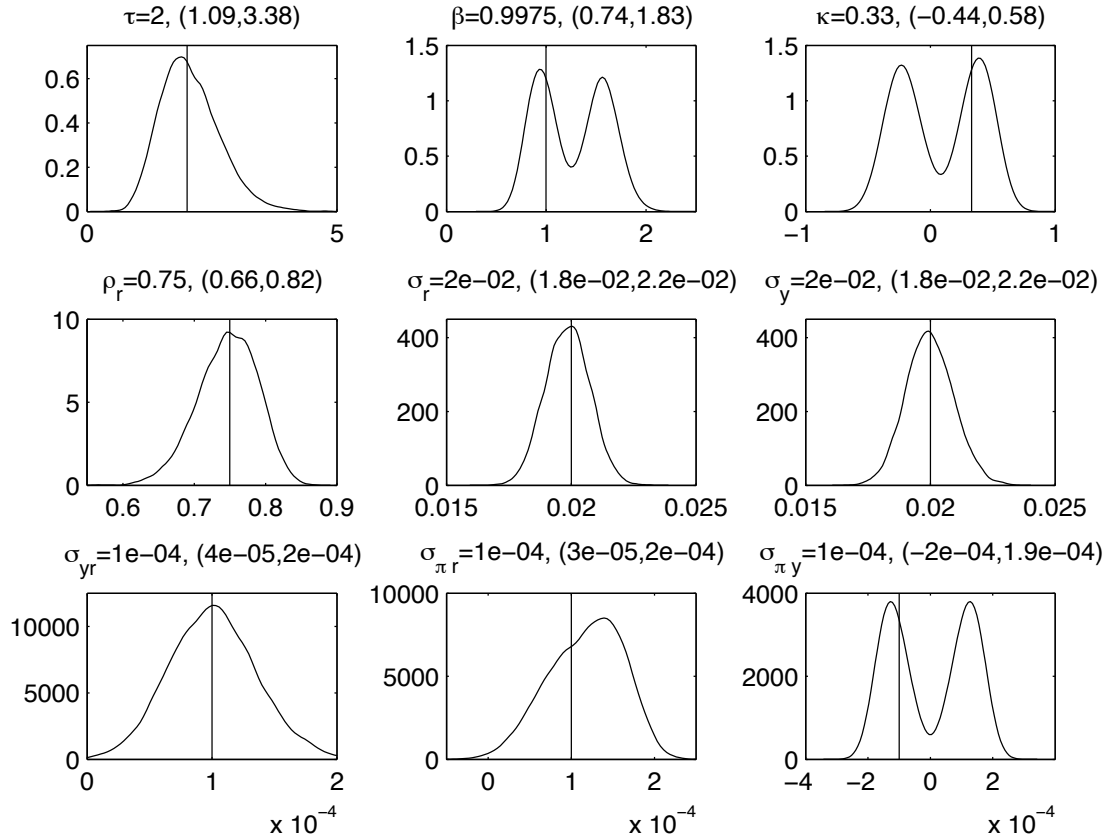
Figure 1.6: Monte Carlo distribution of O Model restricted MLE at ϑ_0 in the real-valued codomain: T=250, N=10,000. Two-sided $(\alpha/2, 1 - \alpha/2)$ confidence interval, $\alpha = 5\%$.

3. If both estimators are in the reals, plot both in the Monte Carlo with equal weight 1.

4. If one is in the reals and one is outside, plot only the one inside with weight of 2^{17} (Note, this option has no effect in the given sample).

The multi-modal character of the likelihood in the parameters β, κ, and $\sigma_{\pi y}$, also tellingly reminiscent of the distribution of weak instrumental variables

---

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

computed by Nelson and Startz (1990), is obvious. Moreover, there are a few other more subtle observations regarding the MLE distribution over the remaining parameters. Firstly, the parameter $\sigma_{\pi r}$ does not seem to have a bimodal distribution, even though it was the fourth structural parameter, besides β, κ, and $\sigma_{\pi y}$, to exhibit observational equivalence in Table 1.4. The reason is that the two modes pile-up immediately next to one another so there appears to be only one rightward-leading peak. Secondly, even single-modal parameters like the coefficient of relative risk aversion $\tau = -\rho_r/(1-\rho_r) \times (1-\phi_{\pi r})/\phi_{yr}$ seems to be biased (leftward). Evidently, even using the longest possible sample of post-war data, there is small sample bias in the MLE.

Yet, the distribution of the maximum likelihood estimator itself is not the only important distribution to consider. As first explicitly shown by Hamilton et al. (2007), normalizations meant to achieve global identification can have unintended consequences on the properties of confidence intervals. Therefore, if the analyst is using a given normalization scheme to identify a given parameter value, it is critical that this identification scheme be embodied in the construction of the Monte Carlo distribution. First, consider again the analyst that uses bounded priors embodied by Θ_0 as the means for identification; specifically, a parameter is identified in Θ_0 if its observationally equivalent point is outside of this space. Recall, this normalization was a successful identifying restriction for ϑ_0 in Table 1.4. If two estimators are both inside of or both outside of Θ_0, there is no reasonable basis to distinguish between the two. Rather, there is only a unique estimator when just one is in Θ_0. The distribution of this estimator is given in Figure 1.7. In this case, the distributions are close to the original MLE, but now the second peaks for β, κ, and $\sigma_{\pi y}$ have shrunk slightly, reflecting the analyst's ability to distinguish between two points when one is outside of Θ_0. Note the standard errors for both β and κ have now changed, indicating a modest but realized affect of the identifying restrictions on the sizes of the two modes. Pseudo-code for Figure 1.7 follows.

1. For draw $\widehat{\Pi}_{\mathtt{i}}$, compute both estimators $\widehat{\vartheta}_{\mathtt{i}}$ and $\widehat{\vartheta}_{\mathtt{i}}^*$ by plugging in to g^{-1}.

2. If both are outside of reals, the estimator is said to not exist for the given draw (22% of utilized sample).
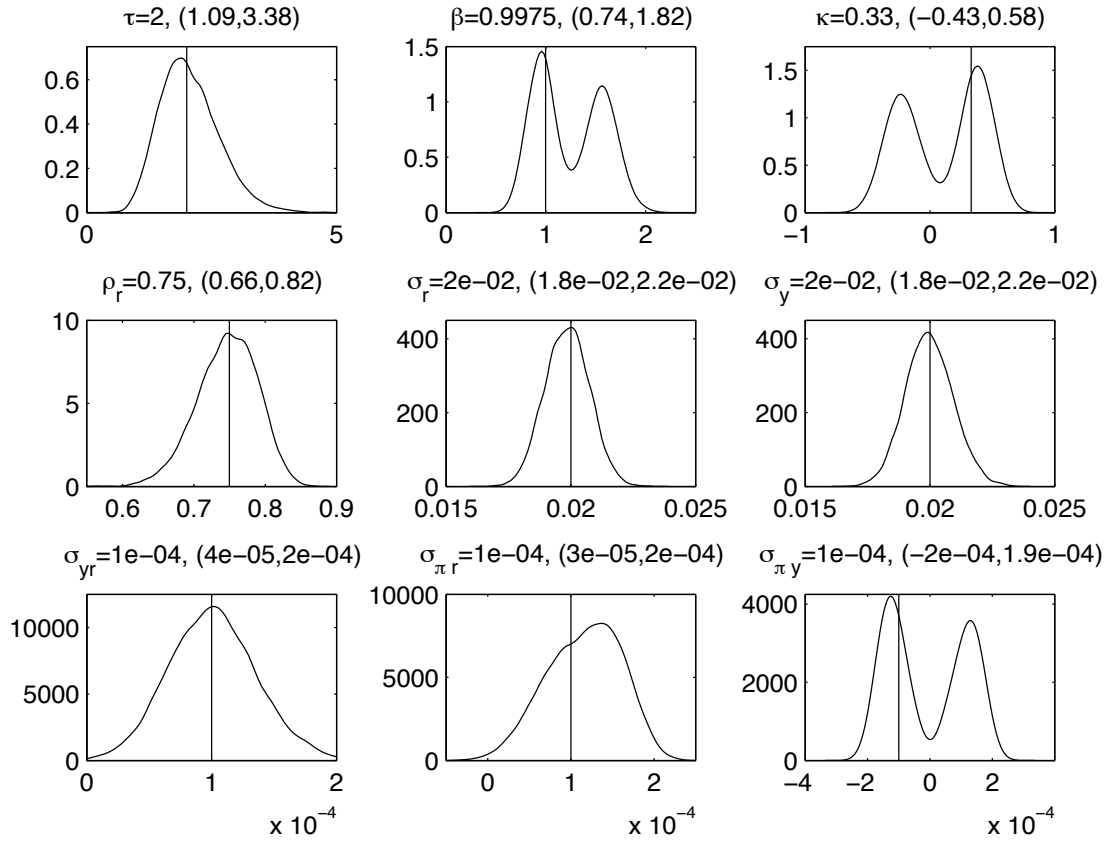
Figure 1.7: Distribution of simplified model MLE at ϑ_0 when the bounds of Θ_0 are used as the identifying restrictions: T=250, N=10,000. Two-sided $(\alpha/2, 1 - \alpha/2)$ confidence interval, $\alpha = 5\%$.

3. If both estimators are in the reals

   (a) If both are inside of Θ_0, the analyst can not distinguish between the two, so plot both in Monte Carlo with equal weight 1.

   (b) If both are outside of Θ_0, the analyst can not distinguish between the two, so plot both in Monte Carlo with equal weight 1.

   (c) If one is in Θ_0 and one is outside, plot only the one inside, with weight 2.

4. If one is in the reals and one is outside, plot only the one inside with weight of 2 (Note, this option has no effect in the given sample).

Now, consider an analyst confronted with a point like ϑ_1 in Table 1.5 for which the normalization that $\vartheta \in \Theta_0$ is not a successful identifying restriction. As was suggested, a normalization that works is to choose the value that yields a β closer to 0.9975, but certainly not larger than 1. The distribution in Figure 1.8 now implies confidence intervals which are notably different from the first MLE without identifying restrictions. In particular, the mode corresponding to the nuisance estimator $\widehat{\vartheta}^*$ has been leveled. Thus, any number of identifying restrictions useful, but confidence intervals are directly affected; the plug-in MLE allows the analyst to directly account for this discrepancy. Pseudo-code for Figure 1.8 follows.

1. For draw $\widehat{\Pi}_i$, compute both estimators $\widehat{\vartheta}_i$ and $\widehat{\vartheta}^*_i$ by plugging in to g^{-1}.

2. If both are outside of reals, the estimator is said to not exist for the given draw (22% of utilized sample).

3. If both estimators are in the reals

   (a) If both imply $\widehat{\beta}_i > 1$, the analyst can not distinguish between the two, so plot both in Monte Carlo with equal weight 1.

   (b) If only one implies $\widehat{\beta}_i > 1$ plot the other in Monte Carlo with weight 2.

   (c) If both imply $\widehat{\beta}_i < 1$

      i. If one estimator yields a $\widehat{\beta}_i$ which is closer to 0.9975, plot only that estimator, with weight 2.

      ii. If both estimators yield $\widehat{\beta}_i$ which are equidistant from 0.9975, plot both estimators, with weight 1 each.

4. If one is in the reals and one is outside, plot only the one inside with weight of 2 (Note, this option has no effect in the given sample).

In this section, I have demonstrated the entirety of my approach to global identification, and accounting for the statistical consequences of necessary normalizations, using the simplified model. Next, I consider the full model, which may be studied similarly.
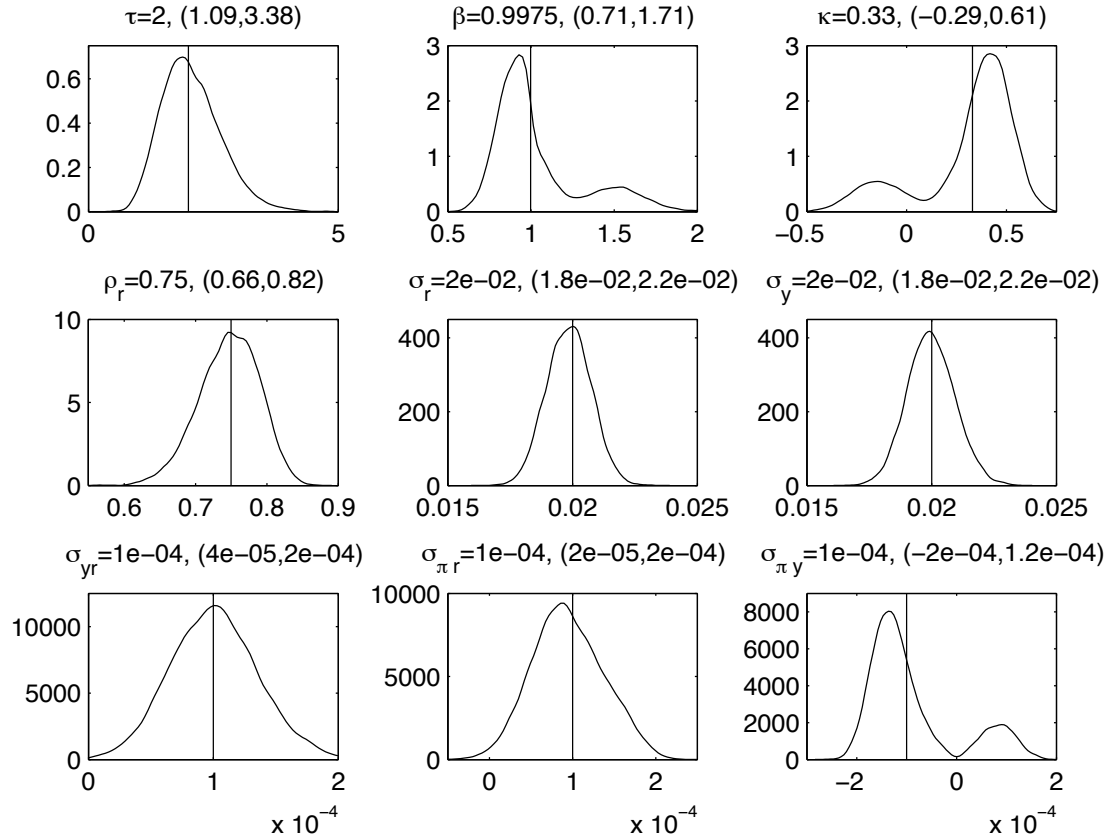
Figure 1.8: Distribution of simplified model MLE with $\beta \approx 0.9975$ identifying restriction at ϑ_0: T=250, N=10,000. Two-sided $(\alpha/2, 1-\alpha/2)$ confidence interval, $\alpha = 5\%$.

5.4 Full Model

Whereas the mapping from θ to Π in the simplified model was complicated, the corresponding mapping for the L model is simply indecipherable from a human perspective; in the computer code for this paper, mathematical expressions for Π in terms of θ are so complicated, they take up pages of text. However, the expressions are closed-form and calculated by the computer, and therefore guaranteed to be correct. A graphical summary is given by Figure 1.9.

As was the case for the simplified model, it is immediately obvious by the necessary order condition $n_\Pi \geq n_\theta$ that the entire vector θ may not be identified. Since there are only $n_\Pi = 13$ reduced form parameters, but $n_\theta = 16$ structural parameters, at least 3 parameters must be set to constants for even a fighting chance that the remaining 13 are conditionally locally or globally identified. Figuring out which are candidates to be set can be done by using the location of $\times$'s in the table, and sequential elimination. The parameters that will be set here are exactly those chosen by Komunjer and Ng (2011) in one of their conditional identification schemes, $\overline{\alpha} = (\overline{\nu}, \overline{\phi}, \overline{\psi}_y)$. These are trained to their values in θ_0. The remaining parameters which may be identified are collected in the 13×1 vector structural parameter

$$\underset{(13 \times 1)}{\vartheta} = (\tau, \beta, \Pi, \psi_\pi, \rho_z, \rho_g, \rho_r, \sigma_z, \sigma_g, \sigma_r, \sigma_{gz}, \sigma_{rz}, \sigma_{rg})'$$

Since $n_\vartheta = 13 = n_\Pi$, if ϑ is identified at a point, it is exactly identified. Indeed, the 13×13 Jacobian $J(\vartheta_0; \overline{\alpha}) = \partial \Pi(\vartheta; \overline{\alpha})/\partial \vartheta'|_{\vartheta=\vartheta_0}$, is full column rank, thus satisfying a necessary and sufficient condition for ϑ to be locally identified at ϑ_0; that the demeaned model is conditionally identified using $\overline{\alpha}$ was previously shown by Komunjer and Ng (2011). But since ϑ is exactly locally identified, and the solution of the model is analytically derived, one can once again hope to derive g^{-1}. Thus, one can observe directly whether g is injective at ϑ_0. The results for one point are given in Figure 1.10, and for many points in.

Also as in the simplified model, even when squared variables such as standard deviations and the steady state of inflation are normalized to positive numbers, there are exactly two ϑ that satisfy $\vartheta = g^{-1}(\Pi; \alpha)$ for any value of Π. Specif-

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

Figure 1.9: An and Schorfheide model functional mapping $g : \theta \to \Pi$.

ically, while the 7 parameters β, ψ_π, ρ_g, ρ_r, σ_g, σ_r, and σ_{rg} are the same in both solutions, all of the 6 parameters τ, Π, ρ_z, σ_z, σ_{gz}, and σ_{rz} differ. But unlike the analysis of the simple model, this result does not actually in itself imply that ϑ is not globally identified in $\mathbb{R}^{13}$ at ϑ_0. In fact, as listed in Figure 1.10, the only realization of ϑ which is observationally equivalent to ϑ_0 is $\vartheta_1 \in \mathbb{C}^{13} \backslash \mathbb{R}^{13}$; in row 3, the value for the steady state of inflation Π at ϑ_1 is $4.8i$.[18] So, even though ϑ is not globally identified in all $\Theta \subset \mathbb{C}^{13}$ at ϑ_0, it is globally identified in the strictly smaller, and more economically meaningful space of all $\Theta \subset \mathbb{R}^{13} \subset \mathbb{C}^{13}$. Impulse responses for ϑ_0 and ϑ_0^* are also given in Figure 1.10, demonstrating that only y_t's response to ε_{gt} is the same for both points. However, while result that ϑ is in fact globally identified in all $\Theta \subset \mathbb{R}^{13}$ at ϑ_0 using $\overline{\alpha}$ is encouraging, this result is specific to ϑ_0, so it is important to assess other points in Θ_0.

Another important point is ϑ_1, listed in Figure 1.11. Although the impulse responses for ϑ_1 and ϑ_1^* appear to be the same at first glance, all but the response of y_t to ε_{gt} in fact differ from point-to-point by very small amounts (all crosshairs are not centered; refer also to the impulse responses in Figure 1.10 for clarity of this point). Thus, even though both points imply similar economic implications, any determination between the two is not technically a normalization as defined in Definition 4, but something stronger. Furthermore, notice that ϑ_1 and ϑ_1^* are both contained in Θ_0. Therefore, these bounds alone may not be used as an identifying restriction as they may be to distinguish between ϑ_0 and ϑ_0^*. An identifying restriction which would work is to simply select the estimator that yields a τ closer to 2, in this case, ϑ_1. Much as the β criterion for the O model reflected tight priors on 0.9975, this similarly reflects the common tight prior on $\tau = 2$.

As was the case for the simplified model, it is of interest to understand how many points are like ϑ_0 versus ϑ_1 in the sense of which identification scheme is successful. Local and global identification is assessed on a uniform grid of Θ in Table 1.7. Once again, ϑ is locally identified for all draws, but now is also globally identified in Θ_0 for nearly all draws. In addition, ϑ is globally identified in all of

---

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

		$g(x,\overline{\alpha})$		ϑ_0	ϑ_0^*
1	ϕ_{rr}	0.79	$\boldsymbol{\tau}$	2	**-45.4**
2	ϕ_{yr}	0.19	β	0.9975	0.9975
3	$\phi_{\pi r}$	0.12	$\boldsymbol{\Pi}$	1.008	**4.8i**
4	ρ_g	0.95	ψ_π	1.5	1.5
5	$\phi_{r\pi}$	0.25	$\boldsymbol{\rho_z}$	0.9	**0.51**
6	$\phi_{y\pi}$	-0.46	ρ_g	0.95	0.95
7	$\phi_{\pi\pi}$	0.62	ρ_r	0.75	0.75
8	ω_r	2e-2	$\boldsymbol{\sigma_z}$	3e-2	**1.85**
9	ω_y	5e-4	σ_g	6e-2	6e-2
10	ω_π	7e-4	σ_r	2e-2	2e-2
11	ω_{yr}	8e-2	$\boldsymbol{\sigma_{gz}}$	1e-4	**-1e-2**
12	$\omega_{\pi r}$	2e-3	$\boldsymbol{\sigma_{rz}}$	1e-4	**1e-2**
13	$\omega_{\pi y}$	4e-2	σ_{rg}	-1e-4	-1e-4
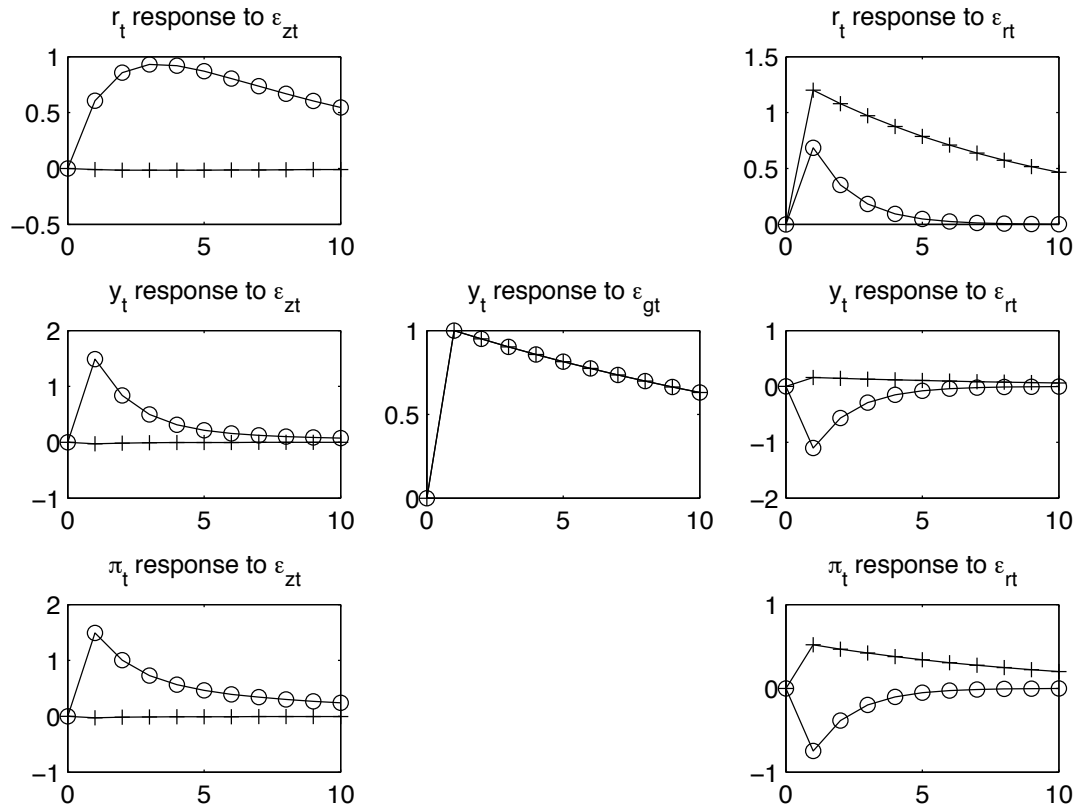
Figure 1.10: Full model without means observational equivalence in $\mathbb{C}^{13}$ but not Θ_0 (see row 3), and impulse-responses: ϑ_0 ($\circ$) vs ϑ_0^* (+). $\Pi = g(x;\overline{\alpha})$; $x = \vartheta_0$ or ϑ_0^*. Missing impulse-responses correspond to restrictions on $D(\theta)$, see ABCD representation.

		$g(x,\overline{\alpha})$		ϑ_1	ϑ_1^*
1	ϕ_{rr}	0.65	τ	1.99	**2.09**
2	ϕ_{yr}	-0.26	β	0.9839	0.9839
3	$\phi_{\pi r}$	-0.14	Π	1.004	**1.029**
4	ρ_g	0.62	ψ_π	1.21	1.21
5	$\phi_{r\pi}$	0.16	ρ_z	0.504	**0.499**
6	$\phi_{y\pi}$	-0.45	ρ_g	0.62	0.62
7	$\phi_{\pi\pi}$	0.35	ρ_r	0.70	0.70
8	ω_r	0.42	σ_z	0.16	**0.12**
9	ω_y	0.97	σ_g	0.45	0.45
10	ω_π	0.37	σ_r	0.57	0.57
11	ω_{yr}	-0.39	σ_{gz}	-0.05	**-0.03**
12	$\omega_{\pi r}$	-0.16	σ_{rz}	0.09	**0.06**
13	$\omega_{\pi y}$	0.35	σ_{rg}	-0.22	-0.22

Figure 1.11: Full model without means observational equivalence in in Θ_0, and impulse-responses: ϑ_1 ($\circ$) vs ϑ_1^* ($+$). $\Pi = g(x;\overline{\alpha})$; $x = \vartheta_1$ or ϑ_1^*.

Table 1.7: Identification of ϑ in full model without means: 100,000 points drawn from Θ_0.

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

$\mathbb{R}^{13}$ in roughly 6/10 draws. So, even though ϑ_0 is part of this majority, there is evidently a large portion of draws at which ϑ is not globally identified in all reals. Yet, again, the parameter space Θ_0 may be used as a reasonable basis to eliminate many of the 40% of points not identified in all of the reals. In fact, only 0.05% of draws yield an observationally equivalent point also in Θ_0. In other words, points like ϑ_1 exist, but are relatively rare compared to points like ϑ_0.

Finally, Monte Carlo distributions are computed. The typical MLE is presented in Appendix A, Figure A.1 and the $\tau \approx 2$ identification criterion is enforced in Figure A.2. In this case, the latter identifying restriction has the added bonus of curtailing a significant fat and skewed tail for $\widehat{\tau}$ observed in Figure A.1. Also, note that the distribution under the Θ_0-based identification scheme, corresponding to Figure 1.7 for the simplified model, is exactly the same as Figure A.1, since there are zero instances in which one estimator is in Θ_0 and one is out.

Finally, with the analysis of the An and Schorfheide model complete, I use the lessons just learned to show how identification of the well-known Smets and Wouters (2007) model may be considered in similar fashion, and conclude.

6   The Smets and Wouters Model

An important question is how to apply these results to a medium-scale DSGE model, such as the Smets and Wouters model. In Appendix D, I prove that the Smets and Wouters model has 7×1 VARMA(3,2) representation

$$Y_t = \sum_{i=1}^{3} \Phi_i Y_{t-i} + U_t + \sum_{j=1}^{2} U_{t-j} \tag{1.23}$$

for which we define

$$\Omega = E\left(U_t + \sum_{j=1}^{2} U_{t-j}\right)\left(U_t + \sum_{j=1}^{2} U_{t-j}\right)'$$

Collect these parameters in the vector

$$\Pi = \left[(\text{vec}\Phi_1)' \quad (\text{vec}\Phi_2)' \quad (\text{vec}\Phi_3)' \quad (\text{vech}\Omega)'\right]' \tag{1.24}$$

Π is not a reduced form parameter, but is trivially identifiable. Therefore, one may create a GMM estimator on the basis of Π, and compute small sample confidence intervals similarly to the procedure described in this paper. Note, processor parallelization across many CPUs is a simple solution to make the many numericaly searches that must be conducted feasible.

7 Conclusion

In this paper I have developed new, simple, reliable, and computationally efficient tools for empirical analysis of a specific well-known DSGE model. As I have shown, the model is not globally identified, but suitable normalizations based on macroeconomic priors are typically available. Although these normalizations may cause the distribution of the maximum likelihood estimator to be nonstandard, the tools I provide also allow the analyst to compute the small sample distribution of the MLE under these restrictions. While discovering global identification failures in DSGE models is difficult in itself, this paper has also argued that it is of equal importance to account for the statistical implications of a given identification procedure.

Chapter 2

Posteriors of Globally Unidentified DSGE Models Under Prior Independence

Abstract. Dynamic stochastic general equilibrium models are typically not globally identified, meaning there are distinct values of the structural parameters that yield the same value of the likelihood function. When the likelihood mode is not unique, neither is the maximum likelihood estimator, and the economic implications of competing observationally equivalent points may differ. One common presumption is that Bayesian estimation directly addresses this problem, since proper priors allow the economist to formally "choose between" observationally equivalent outcomes. This intuition is incorrect. Economic theory yields independent priors, while observational equivalence is characterized by dependence amongst the structural parameters in the likelihood. Only proper dependent priors necessarily partition observationally equivalent outcomes a-posteriori, but such priors are not typically available on the basis of theory alone.

1 Introduction

Consider a DSGE model with likelihood function $\mathcal{L}(Y|\theta)$. $Y = (Y_1', \ldots, Y_T')'$ is a $T n_y \times 1$ dataset of T observations of the $n_Y \times 1$ observables Y_t. $\theta = (\theta_1, \ldots, \theta_{n_\theta})'$

is an $n_\theta \times 1$ vector of structural parameters. The structural parameters θ are not *globally identified* at the point θ_0 if there exists another values of the parameter, $\theta_0^* \neq \theta_0$, that yields the same value of likelihood regardless of data sample, $\mathcal{L}(Y|\theta_0^*) = \mathcal{L}(Y|\theta_0) \, \forall \, Y \in \mathbb{R}^{Tn_Y \times 1}$. In this case, θ_0 and θ_0^* are called *observationally equivalent.*

Recently, the issue of global identification in DSGE models has been studied by Qu and Tkachenko (2013), Kocięcki and Kolasa (2013), and the first chapter of this dissertation. In the first chapter, the main concern is of how the analyst should proceed when faced with the problem. While either admissibility or sign restrictions are useful in distinguishing between observationally equivalent values, it is also shown to be essential to account for the identifying restriction in the computation of confidence intervals.

A common question that is voiced with regards to estimating unidentified DSGE models is of how Bayesian estimators fare. The reasons for optimism are twofold. First, identification failure does not affect the ability to calculate posteriors. Second, since Bayesian estimation allows the economist to specify priors, it is commonly presumed that this will allay the identification problem. Intuitively, proper priors may result in one posterior mode while the likelihood has two. The posterior mode is the main object of interest for computing statistics, like impulse responses, in Smets and Wouters (2007) and other important studies.

While the properties of Bayesian estimators under failure of local identification have been studied by Koop et al. (2013), the implications of global identification failure are less understood. One exception to this rule is the intuition provided in Herbst and Schorfheide (2014). As they explain, global identification failure leads to multimodal posteriors that are reflective of the unusual contours of the likelihood function. For instance, in a model for which there are two observationally equivalent points that maximize the likelihood function, one might reasonably expect there to be two modes of the posterior. The good news, however, is that proper priors often "fix" the identification problem, because the mode of the likelihood which is more heavily weighted by the prior results in a strictly higher mode of the posterior.

The question, however, is whether common priors for DSGE models typically allow the economist to distinguish between observationally equivalent points. In this paper, I will show this to not be the case. The main reason for this finding is that DSGE priors are typically independent; for example, theory suggests that the coefficient of relative risk aversion τ is somewhere between 0.5 and 3.5, and the discount factor β is less than 1, but the joint distribution of the priors is not obvious. Therefore, two points that are observationally equivalent, but for which the analyst only has independent priors, might be hard to disentangle. Joint distributions are therefore necessary to distinguish observational equivalent points a-posteriori, but some priors are hard to determine.

First, I describe the commonly presumed way that priors serve as identifying restrictions. Then, I consider a small-scale DSGE model known to be not globally identified. I show that independent priors do not allow the economist to differentiate between competing observationally equivalent outcomes.

2 Priors as Identifying Restrictions

Most Bayesian analyses of DSGE models are concerned with obtaining the mode of the posterior (See Smets and Wouters (2007), for example). This statistic is used to compute impulse responses, and other intuitive statistics. Therefore, to study the relationship between Bayesian and Classical estimators, it is convenient to consider the maximum a-posteriori estimator (MAP), which is directly related to the maximum likelihood estimator (MLE). Defining the log-likelihood $\ell(Y|\theta) = \ln \mathcal{L}(Y|\theta)$,

$$\widehat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} \ell(Y|\theta) \tag{2.1}$$

while the MAP is written

$$\widehat{\theta}_{MAP} = \arg\max_{\theta \in \Theta} \pi(\theta|Y)$$

where the posterior is defined by

$$\pi(\theta|Y) = \frac{\mathcal{L}(Y|\theta)p(\theta)}{p(Y)}$$

Since the marginal distribution $p(Y)$ is independent of θ, the MAP is equivalently the value of the parameters which maximizes $\mathcal{L}(Y|\theta)p(\theta)$. Furthermore, priors for DSGE models are typically independent. Therefore, $p(\theta)$ may be written

$$p(\theta) = \prod_{i=1}^{n_\theta} p_i(\theta_i)$$

where $p_i(\theta_i)$ is the independent prior for parameter θ_i. Then the MAP may also be written

$$\widehat{\theta}_{MAP} = \arg\max_{\theta\in\Theta} \ell(Y|\theta) + \sum_{i=1}^{n_\theta} \ln p_i(\theta_i) \tag{2.2}$$

The MAP is only necessarily equivalent to the MLE when the priors are uniform. Define the admissible parameter space by

$$\Theta = \{\theta : \underline{\theta}_1 \le \theta_1 \le \overline{\theta}_1, \ldots, \underline{\theta}_{n_\theta} \le \theta_{n_\theta} \le \overline{\theta}_{n_\theta}\}$$

Then uniform prior for each parameter is

$$\mathcal{U}(\theta_i; \underline{\theta}_i, \overline{\theta}_i) = \begin{cases} \frac{1}{\overline{\theta}_i - \underline{\theta}_i} & \text{for } \underline{\theta}_i \le \theta_i \le \overline{\theta}_i \\ 0 & \text{otherwise.} \end{cases}$$

and the MAP under uniform priors is

$$\widehat{\theta}_{MAP} = \arg\max_{\theta\in\Theta} \ell(Y|\theta) - \sum_{i=1}^{n_\theta} \ln(\overline{\theta}_i - \underline{\theta}_i) = \arg\max_{\theta\in\Theta} \ell(Y|\theta) \equiv \widehat{\theta}_{MLE} \tag{2.3}$$

However, if priors are not uniform, the density of the prior depends on the realization of θ and the first-order optimality equation for the MAP is

$$\text{FOC: } \underbrace{0_{1 \times n_\theta} = \frac{\partial \ell(Y|\theta)}{\partial \theta'}\bigg|_{\theta=\widehat{\theta}_{MAP}}}_{\text{MLE FOC}} + \underbrace{\sum_{i=1}^{n_\theta} \iota_i' \left(\frac{1}{p_i(\theta_i)}\bigg|_{\theta=\widehat{\theta}_{MAP}} \times \frac{\partial p_i(\theta_i)}{\partial \theta_i}\bigg|_{\theta=\widehat{\theta}_{MAP}} \right)}_{\text{Additional } 1 \times n_\theta \text{ term}} \quad (2.4)$$

Where ι_i is an $n_\theta \times 1$ vector with a 1 in position i and zeros elsewhere.

One instance in which non-uniform priors might be useful is if there are multiple modes of the likelihood surface. Recall, observational equivalence is defined by

$$\mathcal{L}(Y|\theta) = \mathcal{L}(Y|\theta^*) \text{ for } \theta \neq \theta^* \ \forall \ Y \in \mathbb{R}^{Tn_Y \times 1}$$

When θ and θ^* maximize the likelihood function, the MLE is not unique. However, since the MAP differs from the MLE with non-uniform priors, it is of course possible that there is still one global mode of the posterior. This statistic is appealing because it might be used to "choose between" observationally equivalent points.

I now give an example where the structural parameters are not identified, but proper independent priors allow the economist to select a unique mode of the posterior.

2.1 Example

Consider the atheoretic MA(1) process

$$y_t = \theta_1 \varepsilon_{1t} + \theta_2 \varepsilon_{2t} \quad (2.5)$$

where $Y = (y_1, \ldots, y_T)'$ is a $T \times 1$ vector of observations and ε_{1t} and ε_{2t} are iid Gaussian, $\varepsilon_{1t} \sim \mathcal{N}(0,1)$ and $\varepsilon_{2t} \sim \mathcal{N}(0,1)$. The likelihood for this model is written

$$\mathcal{L}(Y|\theta) = \frac{1}{(2\pi)^{T/2}(\theta_1^2 + \theta_2^2)^{T/2}} \exp\left\{ -\frac{1}{2} \frac{1}{\theta_1^2 + \theta_2^2} \sum_{t=1}^{T} y_t^2 \right\}$$

where $\theta = (\theta_1, \theta_2)'$ and $\Theta = \mathbb{R}^2$. The log-likelihood is proportional to

$$\ell(Y|\theta) \propto -\frac{T}{2} \ln(\theta_1^2 + \theta_2^2) - \frac{1}{2} \frac{1}{\theta_1^2 + \theta_2^2} \sum_{t=1}^{T} y_t^2$$

Therefore, the MLE first order condition is seen to be

$$\widehat{\theta}_{1MLE}^2 + \widehat{\theta}_{2MLE}^2 = \frac{1}{T} \sum_{t=1}^{T} y_t^2$$

Clearly, neither $\widehat{\theta}_{1MLE}$ nor $\widehat{\theta}_{2MLE}$ is individually identifiable; the contours of the likelihood, with a flat surface corresponding to this identification problem are depicted in Figure 2.1 Panel A.

Consider now the Bayesian analyst who uses priors $\theta_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\theta_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ to form the posterior. The MAP is defined by

$$\widehat{\theta}_{MAP} = \arg\max_{\theta \in \Theta} \left\{ -\frac{T}{2} \ln(\theta_1^2 + \theta_2^2) - \frac{1}{2} \frac{1}{\theta_1^2 + \theta_2^2} \sum_{t=1}^{T} y_t^2 - \frac{(\theta_1 - \mu_1)^2}{2\sigma_1^2} - \frac{(\theta_2 - \mu_2)^2}{2\sigma_2^2} \right\}$$

For which the first order conditions are

$$\left(\sum_{t=1}^{T} y_t^2 \right) \left(\frac{1}{\widehat{\theta}_{1MAP}^2 + \widehat{\theta}_{2MAP}^2} \right)^2 - T \left(\frac{1}{\widehat{\theta}_{1MAP}^2 + \widehat{\theta}_{2MAP}^2} \right) + \left(\frac{\mu_i}{\sigma_i^2} \frac{1}{\widehat{\theta}_{iMAP}} - \frac{1}{\sigma_i^2} \right) = 0$$

for $i = 1, 2$. This set of (fifth degree) polynomials in general has a multitude of solutions. For example, if $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 0.1$, then the one FOC is

$$\widehat{\theta}_{1MAP}^2 + \widehat{\theta}_{2MAP}^2 = \frac{T \pm \sqrt{T^2 + 0.4 \sum_{t=1}^{T} y_t^2}}{-0.2}$$

which again has infinitely many solutions. A depiction of the posterior under these priors is given in Figure 2.1 Panel D; notice the exact correspondence with the likelihood.

However, it is possible for the posterior mode to be unique, even given this identification problem. Examples of priors that accomplish this are given in Figure
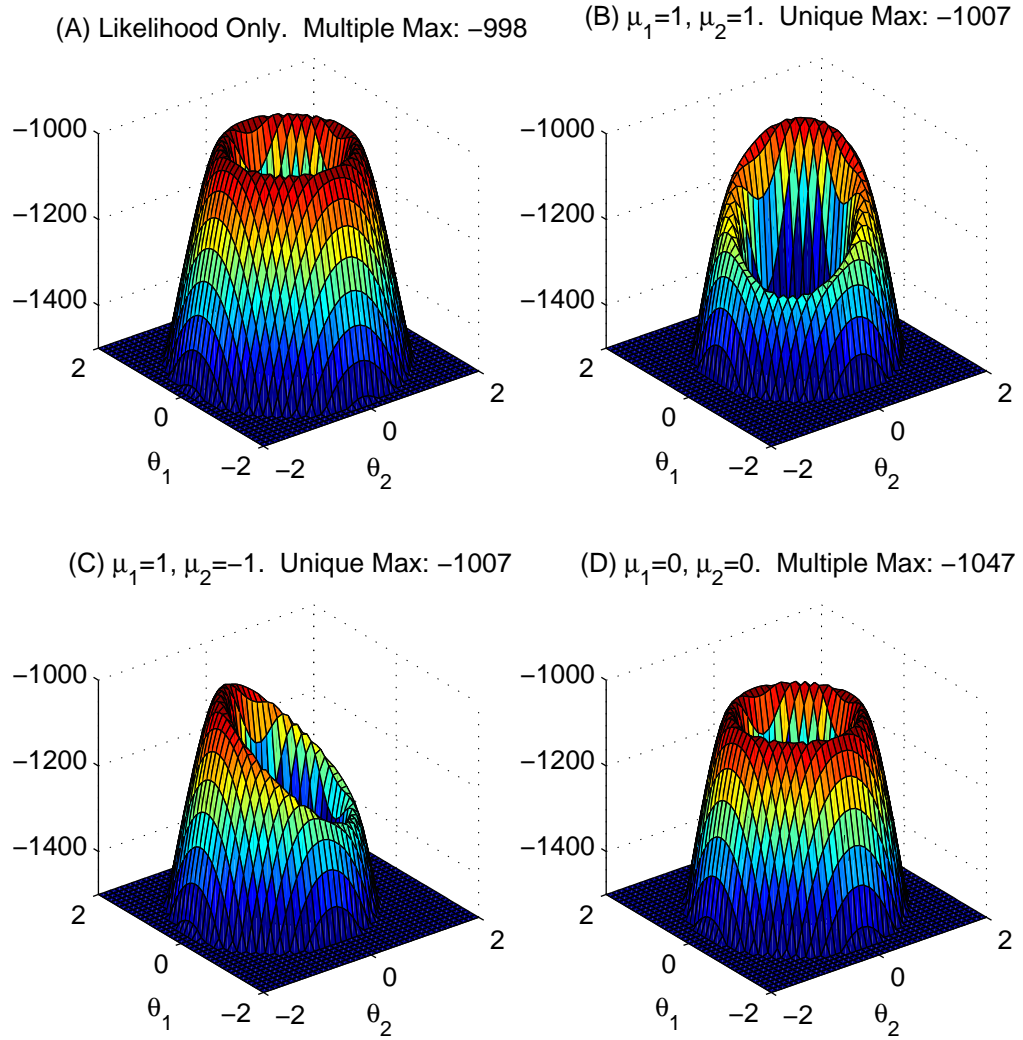
Figure 2.1: Likelihood and posterior contours. (A) Multi-modal likelihood function. (B) Single-modal posterior distribution under priors $\mu_1 \sim \mathcal{N}(1, 0.1)$ and $\theta_2 \sim \mathcal{N}(1, 0.1)$ yields one MAP near (1,1). (C) Single-modal posterior distribution under priors $\mu_1 \sim \mathcal{N}(1, 0.1)$ and $\mu_2 \sim \mathcal{N}(1, 0.1)$ yields one MAP near (1,1). (D) Multi-modal posterior under priors $\mu_1 \sim \mathcal{N}(0, 0.1)$ and $\mu_2 \sim \mathcal{N}(0, 0.1)$.

Table 2.1: Observational equivalence in An and Schorfheide (2007) model.

	$\widehat{\theta}_{MLE}$	$\widehat{\theta}^*_{MLE}$
$\boldsymbol{\tau}$	1.22	**2.02**
β	1.86	1.86
$\boldsymbol{\Pi}$	1.19	**1.54**
ψ_π	1.84	1.84
$\boldsymbol{\rho_z}$	0.9	**0.42**
ρ_g	0.94	0.94
ρ_r	0.75	0.75
$\boldsymbol{\sigma_z}$	0.008	**0.04**
σ_g	6e-2	6e-2
σ_r	2e-2	2e-2
$\boldsymbol{\sigma_{gz}}$	1e-4	**-1e-2**
$\boldsymbol{\sigma_{rz}}$	1e-4	**1e-2**
σ_{rg}	-1e-4	-1e-4

2.1 Panels B and C. Notice, in both cases, the location of priors helps "choose" which of the observationally equivalent values is most in-line with the economist's priors. The MAP corresponds precisely to the prior means.

3 Parameter Dependence, Prior Independence

The above example suggests that priors may be used to "choose" amongst observationally equivalent points. Now, let us consider the DSGE model presented in An and Schorfheide (2007). As proven in the first chapter of this dissertation, this model is not globally identifiable, and any value of the structural parameters has exactly one observationally equivalent value. This means that the maximum likelihood estimator is not unique, and admissibility or sign restrictions must be used to differentiate between the two.

First, using the procedure developed in the first chapter, I create a data set using data generating value θ_0 and obtain two MLEs. These are given in Table 2.1. The prominent difference between the two points is that the CRRA τ is low in $\widehat{\theta}_{MLE}$ and high in $\widehat{\theta}^*_{MLE}$ while ρ_z is high in $\widehat{\theta}_{MLE}$ and low in $\widehat{\theta}^*_{MLE}$.

Since the MLE is not unique, let us alternatively consider the MAP. A typical set of independent priors for the parameters in this model is given in Table

Table 2.2: Priors for An and Schorfheide (2007).

Parameter	Prior	Mean	Std Dev
τ	$\mathcal{G}$	2	0.25
$100 \times (1/\beta - 1)$	$\mathcal{G}$	$0.2506 = 100 \times (1/0.9975 - 1)$	0.25
$1000 \times (\Pi - 1)$	$\mathcal{G}$	$8 = 1000 \times (1.008 - 1)$	6
ψ_π	$\mathcal{G}$	1.5	1
ρ_z	$\mathcal{B}$	0.5	0.25
ρ_g	$\mathcal{B}$	0.5	0.25
ρ_r	$\mathcal{B}$	0.5	0.25
σ_z	$\mathcal{IG}$	0.05	0.01
σ_g	$\mathcal{IG}$	0.05	0.01
σ_r	$\mathcal{IG}$	0.05	0.01
$\rho_{gz} = \frac{\sigma_{gz}}{\sigma_g \sigma_z}$	$\mathcal{N}$	0	0.25
$\rho_{rz} = \frac{\sigma_{rz}}{\sigma_r \sigma_z}$	$\mathcal{N}$	0	0.25
$\rho_{rg} = \frac{\sigma_{rg}}{\sigma_r \sigma_g}$	$\mathcal{N}$	0	0.25

2.2. I verify using Andreasen (2010)'s genetic search algorithm that the posterior has only one global mode on the support of the prior. However, if the posterior has the quality that the prior has "chosen between" observationally equivalent points, it should be the case that an analyst with drastically different priors for any of the parameters that vary between observationally equivalent points will obtain a different posterior. Is this the case?

In order to test this hypothesis, I vary the mean of τ's prior between 0.5 and 5 in increments of 0.1, and use Andreasen's algorithm to find the posterior mode in each case. The results of this experiment are given in Appendix B Figure B.1. Clearly, shifting the mean of τ's prior does not result in a discrete shift from one mode to another, as might be expected, but only incremental changes in the location of τ's posterior. In Appendix B Figure B.2, I show that changing the means of the priors for τ and ρ_z at the same time seems to only affect the posterior mode for each parameter individually. One possible explanation for this is that the posterior is discontinuous; since DSGE models are not variation-free, there are large portions of the parameter space in which the likelihood is not defined.

The implication of the above experiment is that two economists with varying priors about independent parameters will not compute posterior modes that are representative about the difference between $\widehat{\theta}_{MLE}$ and $\widehat{\theta}^*_{MLE}$. Rather, only

dependent priors about the joint distribution about the parameters which differ between $\widehat{\theta}_{MLE}$ and $\widehat{\theta}_{MLE}^*$ will cause the posterior mode to differ in all parameters.

4   Conclusion

In this paper, I have disspelled the commonly held notion that standard DSGE priors may help the analyst "choose between" between observationally equivalent economic stories from the data. Because priors are independent, but observational equivalence emerges from nonlinear dependence amongst parameters, varying one prior at a time only necessarily effects the placement of the posterior that individual parameter. Only substantially different joint priors will necessarily affect the placement of the posterior substantially, but such priors are typically hard to come by.

Chapter 3

Local Identification of Nonlinear DSGE Models

Abstract. While rank and order conditions for the identification of linearized DSGE models have recently been introduced, no formal results exist on the identifiability of nonlinear models. In this paper, I show how to represent the nonlinear pruned state space system derived in Andreasen et al. (2014) in minimal linear state space representation. I use this reparameterization to apply the rank and order conditions derived in Komunjer and Ng (2011), originally intended for linearized models, to the nonlinear case. I confirm An and Schorfheide (2007)'s intuition that the elasticity of demand and price stickiness are identifiable in a nonlinear approximation of their model, but not linear.

1   Introduction

In a well-known paper, An and Schorfheide (2007) study a canonical DSGE specification. In studying the identifiability of this model, they make two claims. First, they assert that in a linear approximation of their model, the elasticity of demand and price stickiness are not separately identifiable (p. 122), and that steady state government spending is unidentified (p. 164). Second, they posit that all three parameters are potentially identifiable simply by using a nonlinear approximation.

It was not until after An and Schorfheide's paper, with the contributions of Canova and Sala (2009), Iskrev (2010), Komunjer and Ng (2011), and Qu and Tkachenko (2012), that identification of linearized DSGE models was studied completely rigorously. In particular, Komunjer and Ng have provided generally applicable rank and order conditions, and a thorough analysis of the An and Schorfheide model in particular. As Komunjer and Ng show, all of the elasticity of demand, price stickiness, and steady state government spending must be set to constants for the complement set to be locally identified in a linear approximation, thus confirming the first half of An and Schorfheide's intuition.

Yet, because rank and order conditions for nonlinear DSGE models have not been provided, the second half of An and Schorfheide's claim remains unverified; are these three parameters identifiable, simply by utilizing a nonlinear solution? Furthermore, can using nonlinear approximations enhance the identifiability of DSGE parameters more generally?

The answer to both questions is yes. In this paper, I provide a methodology for assessing local identification of nonlinear DSGE models. Since determining the identifiability of linear models is already a difficult problem, I approach the problem not by deriving new rank and order conditions, but by repurposing old ones. In the first substantive section of our paper, I show that the class of nonlinear pruned state space systems presented in Andreasen et al. (2014) may be reparameterized to minimal ABCD or AKCΣ "innovations" representation. Since this is immediately the input to the Komunjer and Ng conditions for linearized models, I am able to assess the identifiability of nonlinear models using their results. A central observation I make is that the minimal representation of the linearized model is nested within the minimal representation of the nonlinear model. This nestedness feature allows the Komunjer and Ng rank and order conditions to be satisfied by a nonlinear approximation of a given model, even when they are not in a linearized version of the same model.

In the following, I discuss the reparameterization of pruned nonlinear state space to minimal state space representation. I use the simple model presented in Schmitt-Grohé and Uribe (2004) to demonstrate the approach. Next, I use

this minimal reparameterization to study the identifiability of nonlinear models. I show that An and Schorfheide's predictions for enhanced identifiability in nonlinear approximations were correct.

2 Representation of Nonlinear DSGE Models

Let x_t be a $n_x \times 1$ vector of detrended, predetermined state variables, where $n_x < \infty$. Let y_t be a $n_y \times 1$ vector of detrended but non-predetermined control variables, where $n_y < \infty$. Finally, let θ be an $n_\theta \times 1$ vector of structural parameters which belongs to the set $\Theta \subseteq \mathbb{R}^{n_\theta}$. I consider DSGE models of the form

$$E_t f\left(x_{t+1}, x_t, y_{t+1}, y_t | \theta\right) = 0_{n_x + n_y} \tag{3.1}$$

where $0_{n_x+n_y}$ is an $(n_x + n_y) \times 1$ vector of zeros. As discussed in Schmitt-Grohé and Uribe (2004), the solution of this model may be written as a set of decision rules depending on x_t and a perturbation parameter $\sigma \geq 0$.[1] The decision rule for y_t is the control equation.

$$y_t = g(x_t, \sigma | \theta) \tag{3.2}$$

The state vector x_t's decision rule is called the state equation.

$$x_t = h(x_{t-1}, \sigma | \theta) + \sigma \eta(\theta) u_t \tag{3.3}$$

where u_t is an $n_u \times 1$ vector of exogenous white noise shocks, $u_t \sim WN(0_{n_u \times 1}, I_{n_u})$, which is strictly more general than IID. η is a rectangular matrix with dimension $n_x \times n_u$.[2] For clarity of exposition, I will henceforth interpret the non-

---

[1]The perturbation parameter accounts for precautionary behavior induced by the expected variance of future shocks. In a linearized model, the only aspect of future shocks that affects the agent's decision-making process is their expected value, which is typically zero. However, in a second-order approximation of the same model, the second moments (variances) of future shocks matter, and are always nonzero.

[2]To accomodate a non-identity positive definite covariance matrix $\Sigma_u(\theta)$ for the innovations u_t, the matrix $\eta(\theta)$ may be written as a matrix product $\eta(\theta) = N(\theta) \times L_u(\theta)$, where N is an arbitrary $n_x \times n_u$ matrix, and $L_u(\theta)$ is the Cholesky decomposition of Σ_u.

predetermined variables y_t to be the observable variables in the data.[3]

The functions g and h are almost never known in closed-form. For this reason, Schmitt-Grohé and Uribe proposed approximating their Taylor series expansions using a perturbation algorithm. However, for expansions of order higher than one, it was widely observed that impulse-responses tended to diverge. This result contradicts the initial assumption of steady state, around which the Taylor series approximation is made. In response, Kim et al. (2008) proposed a second step of eliminating ("pruning") certain terms from the series expansion (See also Lombardo and Sutherland (2007)). Pruned state space models yield convergent impulse responses, but are no longer Taylor series of the antecedent microfounded solution. What makes pruned state space models useful is that in conjunction with convergent impulse-responses, the errors of pruned and unpruned state space models are frequently of the same order. This is shown, for example, in Andreasen et al. (2014).

Andreasen et al. also explore pruned state space models' potential for estimation. They do so by showing how to compute second moments in closed form, and applying moment-based estimators, such as GMM. Yet, the identifiability of the structural parameters in these models is not understood, which is a primitive assumption underlying the consistency of such estimators. Identification problems have characterized the estimation of linearized DSGE models, and resolving those issues is not always straightforward (See Canova and Sala (2009)). It is unclear how nonlinear approximations fare versus linear with respect to the identifiability of key macroeconomic parameters.

Next, I review pruned state space representation of second order approximations of DSGE models. The derivation works in two steps, by first approximating the nonlinear solution of the model, and then "pruning" that approximation. The ultimate pruned functional form will provide the foundation for rigorous identification analysis in the subsequent sections of the analysis.

---

[3]This assumption is made without loss of generality only to relieve obfuscation of the main results by excessive matrix algebra operations. ($\rightarrow$) Any elements in x_t for which data is available can be related to y_t by an identity in the function g. ($\leftarrow$) Any variable originally included in y_t that is not observable can be moved to x_t. These points will be clarified in later examples.

2.1 Second Order Approximation

I begin by deriving standard unpruned approximations of the solution. A first-order Taylor series expansion of the state equation (3.3) about the deterministic steady state $(x_t = x_{t-1} = x^*(\theta), \sigma = 0)$ is

$$\widehat{x}_t \approx h_x(\theta)\widehat{x}_{t-1} + \overset{0}{h_\sigma(\theta)\sigma} + \sigma\eta(\theta)u_t \tag{3.4}$$

$\widehat{x} = x_t - x^*$ is the deviation of the states x_t from steady state $x^* = h(x^*, 0|\theta)$ and

$$h_x(\theta) = \left.\frac{\partial h(x_{t-1}, \sigma|\theta)}{\partial x'_{t-1}}\right|_{x_{t-1}=x^*, \sigma=0} \qquad h_\sigma(\theta) = \left.\frac{\partial h(x_{t-1}, \sigma|\theta)}{\partial \sigma'}\right|_{x_{t-1}=x^*, \sigma=0}$$
$$\scriptstyle n_x \times n_x \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad n_x \times 1$$

The observation equation's first-order series expansion is

$$\widehat{y}_t = g_x(\theta)\widehat{x}_t + \overset{0}{g_\sigma(\theta)\sigma} \tag{3.5}$$

$\widehat{y}_t = y_t - y^*(\theta)$ is the deviation of the observables y_t from steady state $y^* = g(x^*, 0|\theta)$ and

$$g_x(\theta) = \left.\frac{\partial g(x_t, \sigma|\theta)}{\partial x'_t}\right|_{x_t=x^*, \sigma=0} \qquad g_\sigma(\theta) = \left.\frac{\partial g(x_t, \sigma|\theta)}{\partial \sigma'}\right|_{x_t=x^*, \sigma=0}$$
$$\scriptstyle n_y \times n_x \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad n_y \times 1$$

That both $h_\sigma = 0$ and $g_\sigma = 0$ is proven formally by Schmitt-Grohé and Uribe. Second-order Taylor series expansions of the state and observation equations are written

$$\widehat{x}_t \approx h_x(\theta)\widehat{x}_{t-1} + \overset{0}{h_\sigma(\theta)\sigma} + \frac{1}{2}H_{xx}(\theta)\widehat{x}_{t-1}^{\otimes 2} + \overset{0}{h_{x\sigma}\widehat{x}_{t-1}\sigma} + \frac{1}{2}h_{\sigma\sigma}(\theta)\sigma^2 + \sigma\eta(\theta)u_t \tag{3.6}$$

$$\widehat{y}_t \approx g_x(\theta)\widehat{x}_t + \overset{0}{g_\sigma(\theta)\sigma} + \frac{1}{2}G_{xx}(\theta)x_t^{\otimes 2} + \overset{0}{g_{x\sigma}\widehat{x}_t\sigma} + \frac{1}{2}g_{\sigma\sigma}(\theta)\sigma^2 \tag{3.7}$$

Details on the functional form of the coefficient matrices appearing in (3.6) and (3.7) are given in Appendix B.1, and that the cross-partials between states and perturbation parameter are zero is proven by Schmitt-Grohé and Uribe. Finally, I

have made use of the convenient shorthand

$$\widehat{x}_t^{\otimes n} = \underbrace{\widehat{x}_t \otimes \ldots \otimes \widehat{x}_t}_{n \text{ times}}$$

for $\otimes$ the Kronecker product (see Abadir and Magnus (2005)).

With this second-order approximation of the state and observations equations in-hand, the next step is to "prune" the expansion. I now discuss pruning.

2.2 The Pruned State Space System: Baseline Case

The second-order approximation (3.6) implies that $\widehat{x}_t$ is only a second order polynomial in the elements of $\widehat{x}_{t-1}$, since it is linear in $\widehat{x}_{t-1}^{\otimes 2}$. However, since $\widehat{x}_{t-1}$ is also linear in $\widehat{x}_{t-2}^{\otimes 2}$, $\widehat{x}_t$ is a function of third and fourth-order terms in period $t-2$, including $\widehat{x}_{t-2}^{\otimes 3}$ and $\widehat{x}_{t-2}^{\otimes 4}$. Inductively, each term $\widehat{x}_t$ is a function of $\lim_{n \to \infty} u_{t-n}^{\otimes n}$. Therefore, a shock to u_t is potentially explosive in its implied dynamics for $\lim_{n \to \infty} \widehat{x}_{t+n}$, and belies the original assumption of steady state. All approximations of the solution of order two or above are prone to similar dynamic inconsistency.

Pruning is a method of augmenting the model to prevent such explosive dynamics, while maintaining the accuracy of the approximation (measured, for example, by Euler equation errors). To motivate it, recall, perturbation is a method that approximates the Taylor series expansion of the solution at one point in time; when the expansion is second order, approximation error is third-order. Second-order pruning is similarly a method that removes terms of order three and above. However, these are terms from the expansion of $\widehat{x}_t$ over time, such as $\widehat{x}_{t-2}^{\otimes 3}$, $\widehat{x}_{t-2}^{\otimes 4}$, and $\widehat{x}_{t-3}^{\otimes 8}$. Since these terms are small, the hypothesis is that removing them will not affect the substantive economic implications of the model. The validity of this claim is justified, for example, by Andreasen et al. (2014).

Andreasen et al. prune as follows: Let $\widehat{x}_t^s$ represent the state vector with a rule of motion corresponding to the second order approximation, (3.6); $s=$"second-order." Let $\widehat{x}_t^f$ correspond to an entirely separate state vector, with a rule of motion corresponding to a first order series, (3.4); $f=$"first-order." Now, in (3.6), replace

$\widehat{x}_t$ with $\widehat{x}_t^f + \widehat{x}_t^s$. This substitution yields the system

$$\widehat{x}_t^f + \widehat{x}_t^s = h_x(\theta)(\widehat{x}_{t-1}^f + \widehat{x}_{t-1}^s) + \frac{1}{2}H_{xx}(\theta)(\widehat{x}_{t-1}^f + \widehat{x}_{t-1}^s)^{\otimes 2} + \frac{1}{2}h_{\sigma\sigma}(\theta)\sigma^2 + \sigma\eta(\theta)u_t \quad (3.8)$$

$$\widehat{x}_t^f = h_x(\theta)\widehat{x}_{t-1}^f + \sigma\eta(\theta)u_t \quad (3.9)$$

$$\widehat{x}_t^s = h_x(\theta)\widehat{x}_{t-1}^s + \frac{1}{2}H_{xx}(\theta)\widehat{x}_{t-1}^{s\otimes 2} + \frac{1}{2}h_{\sigma\sigma}(\theta)\sigma^2 + \sigma\eta(\theta)u_t \quad (3.10)$$

To reduce this system by pruning, first observe that the quadratic terms of interest

$$(\widehat{x}_{t-1}^f + \widehat{x}_{t-1}^s)^{\otimes 2} = \widehat{x}_{t-1}^f \otimes \widehat{x}_{t-1}^f + \widehat{x}_{t-1}^f \otimes \widehat{x}_{t-1}^s + \widehat{x}_{t-1}^s \otimes \widehat{x}_{t-1}^f + \widehat{x}_{t-1}^s \otimes \widehat{x}_{t-1}^s$$

Both $\widehat{x}_{t-1}^f \otimes \widehat{x}_{t-1}^s$ and $\widehat{x}_{t-1}^s \otimes \widehat{x}_{t-1}^f$ are inductively functions of $\widehat{x}_{t-2}^{s\otimes 2} \otimes \widehat{x}_{t-2}^f$. Meanwhile, $\widehat{x}_{t-1}^s \otimes \widehat{x}_{t-1}^s$ is a function of $\widehat{x}_{t-2}^{s\otimes 4}$. Since each of the terms $\widehat{x}_{t-1}^{s\otimes 2} \otimes \widehat{x}_{t-1}^f$ and $\widehat{x}_{t-1}^{s\otimes 4}$ are of order higher than two, they are "pruned" off the expansion. In other words, they are collected into an third-order error. Thus, $(\widehat{x}_t^f + \widehat{x}_t^s)^{\otimes 2} \approx \widehat{x}_t^f \otimes \widehat{x}_t^f$, so

$$\widehat{x}_t^s \approx h_x(\theta)\widehat{x}_{t-1}^s + \frac{1}{2}H_{xx}(\theta)\widehat{x}_{t-1}^{f\otimes 2} + \frac{1}{2}h_{\sigma\sigma}(\theta)\sigma^2 \quad (3.11)$$

Similar operations on the control equation yield

$$\widehat{y}_t^f + \widehat{y}_t^s \approx g_x(\theta)(\widehat{x}_t^f + \widehat{x}_t^s) + \frac{1}{2}G_{xx}(\theta)\widehat{x}_t^{f\otimes 2} + \frac{1}{2}g_{\sigma\sigma}(\theta)\sigma^2 \quad (3.12)$$

Equations (3.11) and (3.12) encapsulate second order pruned state space dynamics. However, it is useful to consider slightly more compact notation.

$$
\begin{bmatrix} \widehat{x}_t^f \\ \widehat{x}_t^s \\ \widehat{x}_t^{f\otimes 2} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2}h_{\sigma\sigma}\sigma^2 \\ \sigma^2\eta^{\otimes 2}\mathrm{vec}(I_{n_u^2}) \end{bmatrix} + \begin{bmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2}H_{xx} \\ 0 & 0 & h_x^{\otimes 2} \end{bmatrix} \times \begin{bmatrix} \widehat{x}_{t-1}^f \\ \widehat{x}_{t-1}^s \\ \widehat{x}_{t-1}^{f\otimes 2} \end{bmatrix}
$$
$$
+ \begin{bmatrix} \sigma\eta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \sigma^2\eta^{\otimes 2} & \sigma(\eta \otimes h_x) & \sigma(h_x \otimes \eta) \end{bmatrix} \times \begin{bmatrix} u_t \\ u_t^{\otimes 2} - \mathrm{vec}(I_{n_u^2}) \\ \mathrm{vec}(\widehat{x}_{t-1}^f u_t') \\ \mathrm{vec}(u_t \widehat{x}_{t-1}^{f'}) \end{bmatrix} \quad (3.13)
$$

The rule of motion for the observables (3.12) may also be rewritten in terms the expanded state vector as

$$\widehat{y}_t^f + \widehat{y}_t^s = \frac{1}{2}g_{\sigma\sigma}(\theta)\sigma^2 + \begin{bmatrix} g_x & g_x & \frac{1}{2}G_{xx} \end{bmatrix} \begin{bmatrix} \widehat{x}_t^f \\ \widehat{x}_t^s \\ \widehat{x}_t^{f\otimes 2} \end{bmatrix} \tag{3.14}$$

Equations (3.13) and (3.14) are the main objects of interest in Andreasen et al.'s analysis, and are called pruned state space representation. The preliminary contribution of this paper will be to show that many elements of this representation are redundant and/or nonminimal. Before returning to this claim, I first consider a class of pruned state space models which will become useful in the following sections.

2.3   Generalized Case: Nonlinearity in Errors and States

The state equation (3.3) implies linear independence between the states x_t and errors u_t. This is problematic if the theory at hand implies nonlinearities between states and errors, a case I will show to be typical. Yet, this can be accommodated by the current set-up. Say that the model implies nonlinearity between the states and some errors having covariance matrix $\Sigma_u(\theta)$ with Cholesky decomposition $L_u(\theta)$. Defining these errors v_t and expanding x_t to $\begin{bmatrix} x_t' & v_{t+1}' \end{bmatrix}'$, the following has the same functional form as (3.3):[4]

$$\begin{bmatrix} x_t \\ v_{t+1} \end{bmatrix} = \begin{bmatrix} h(x_{t-1}, v_t, \sigma|\theta) \\ 0 \end{bmatrix} + \sigma \begin{bmatrix} 0 \\ L_u \end{bmatrix} u_{t+1} \tag{3.15}$$

Given this setup, and using the arguments introduced in the previous section, a first order Taylor approximation of the rule of motion for $\widehat{x}_t$ may be written in

---

[4]Note, I have momentarily advanced the timing convention forward one period from the previous convention in Equation (3.3); (3.15) has the functional form $s_{t+1} = h(s_t, \sigma|\theta) + \sigma\eta u_{t+1}$ for states s_t.

exactly the same functional form as (3.4), given we have defined

$$\underset{n_x \times n_u}{\eta(\theta)} = \left(\left. \frac{\partial h(x_{t-1}, v_t, \sigma|\theta)}{\partial v_t'} \right|_{x_{t-1}=x^*, u_t=0, \sigma=0} \right) \times L_u$$

In contrast, a distinction arises with respect to the second order approximation, due to the nonlinearity between x_{t-1} and v_t. In this case, the second-order accurate expansion of the state equation is

$$\widehat{x}_t \approx h_x(\theta)\widehat{x}_{t-1} + \frac{1}{2}H_{xx}(\theta)\widehat{x}_{t-1}^{\otimes 2} + \frac{1}{2}\sigma H_{xu}(\theta)\mathrm{vec}(\widehat{x}_{t-1}u_t')$$

$$+ \frac{1}{2}\sigma^2 H_{uu}(\theta)u_t^{\otimes 2} + \frac{1}{2}\sigma H_{ux}(\theta)\mathrm{vec}(u_t\widehat{x}_{t-1}') + \frac{1}{2}h_{\sigma\sigma}(\theta)\sigma^2 + \sigma\eta(\theta)u_t \quad (3.16)$$

The functional forms of the coefficient matrices in (3.16) are given in Appendix Section B.2. Mimicking the pruning steps in the previous section, this approximation implies the system

$$
\begin{bmatrix} \widehat{x}_t^f \\ \widehat{x}_t^s \\ \widehat{x}_t^{f\otimes 2} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2}h_{\sigma\sigma}\sigma^2 + \frac{1}{2}\sigma^2 H_{uu}I_{n_u^2} \\ \sigma^2\eta^{\otimes 2}\mathrm{vec}(I_{n_u^2}) \end{bmatrix} + \begin{bmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2}H_{xx} \\ 0 & 0 & h_x^{\otimes 2} \end{bmatrix} \times \begin{bmatrix} \widehat{x}_{t-1}^f \\ \widehat{x}_{t-1}^s \\ \widehat{x}_{t-1}^{f\otimes 2} \end{bmatrix}
$$

$$
+ \begin{bmatrix} \sigma\eta & 0 & 0 & 0 \\ 0 & \frac{1}{2}\sigma^2 H_{uu} & \frac{1}{2}\sigma H_{xu} & \frac{1}{2}\sigma H_{ux} \\ 0 & \sigma^2\eta^{\otimes 2} & \sigma(\eta \otimes h_x) & \sigma(h_x \otimes \eta) \end{bmatrix} \times \begin{bmatrix} u_t \\ u_t^{\otimes 2} - \mathrm{vec}(I_{n_u^2}) \\ \mathrm{vec}(\widehat{x}_{t-1}^f u_t') \\ \mathrm{vec}(u_t\widehat{x}_{t-1}^{f'}) \end{bmatrix} \quad (3.17)
$$

Thus, the pruned state space system (3.17) has the same functional form as (3.13), aside from differences in exclusion restrictions that have arisen from second-order dependence between $\widehat{x}_{t-1}^f$ and u_t. In other words, (3.13) is simply the special case of (3.17) in which H_{uu}, H_{xu}, and H_{ux} are all zero-matrices. Finally, nonlinearities between states and errors in the observation equation function g only need not be considered directly, since any element of y_t may be placed in x_t as well. Therefore, the pruned observation equation approximation for the current case remains (3.14).

While the generalized pruned state space, equations (3.17) and (3.14), is

relatively compact, it is distinct from a purely linear system. Yet, there are in fact many statistical similarities between representations. In the following section, I show that the pruned state space may be reparameterized to minimal state space representation. This insight will be central to identification analysis.

3 Minimal Representation

In this section, first, I show that deviations-from-means of the pruned state space may be reparameterized to ABCD representation. This representation, familiar from Fernández-Villaverde et al. (2007), was previously thought to be applicable only to linear approximations of DSGE models. The ABCD model may also be written in so-called AKCΣ "innovations" representation, which recasts the system in terms of optimal linear forecasts and forecast errors. Second, exploiting ABCD representation, I show that the pruned state space model is not minimal. I show how to condense the model to satisfy minimality, setting the stage for identification analysis. Finally, I show how to carry out the aforementioned reparameterization in an example.

3.1 The Pruned State Space is Nonminimal ABCD

As a preliminary step, I show how to directly reparameterize the generalized pruned state space model equations (3.17) and (3.14) to ABCD and AKCΣ form. To do so, I make three general assumptions regarding the pruned state space solution.

Assumption 1. *The modulus of all eigenvalues of h_x are less than one.*

Assumption 2. *The zeros in h_x, H_{xx}, H_{xu}, H_{ux}, and H_{uu} do not vary over $\theta \in \Theta$.*

Assumption 3. *The fourth moments of u_t are finite.*

Assumptions 1 and 3 are also made by Andreasen et al. Assumption 2 is typically implicit of DSGE analysis. Using these assumptions, I derive ABCD representation in Appendix C in three simple steps. The conclusion of these operations may be summarized with a concise proposition.

Proposition 1. *Under Assumptions 1, 2, and 3, the pruned state space (3.17)
and (3.14) may be written in terms of deviations-from-means as an ABCD model*

$$
\begin{aligned}
X_t &= A(\theta) \ X_{t-1} \ + \ B(\theta) \ \varepsilon_t \\
Y_t &= C(\theta) \ X_{t-1} \ + \ D(\theta) \ \varepsilon_t
\end{aligned}
\tag{3.18}
$$

where the dimensions of the variables are denoted n_X, n_Y, and n_ε.

The ABCD representation of the model involves a nontrivial transformation. In terms of the variables and parameters defined thus far, the elements of ABCD representation are written as follows. First, the state vector X_t is defined as

$$
X_t = M \times \widehat{Z}_t \ \text{ for } \ \widehat{Z}_t = Z_t - E(Z_t|\theta)
$$

with

$$
Z_t = \begin{bmatrix} \widehat{x}_t^f \\ \widehat{x}_t^s \\ D_{n_x}^+ \widehat{x}_t^{f \otimes 2} \end{bmatrix} \ \text{ and } \ E(Z_t|\theta) = (I_{n_Z} - P(\theta))^{-1} J(\theta)
$$

for $D_{n_x}^+$ the Moore-Penrose pseudo inverse of the n_x-dimensional duplication matrix D_{n_x}, and K and P ancillary parameters defined by

$$
J(\theta) = \begin{bmatrix} 0 \\ \frac{1}{2} h_{\sigma\sigma}\sigma^2 + \frac{1}{2}\sigma^2 H_{uu} I_{n_u^2} \\ \sigma^2 D_{n_x}^+ \eta^{\otimes 2} \mathrm{vec}(I_{n_u^2}) \end{bmatrix}_{n_Z \times 1} \qquad P(\theta) = \begin{bmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2} H_{xx} D_{n_x} \\ 0 & 0 & D_{n_x}^+ h_x^{\otimes 2} D_{n_x} \end{bmatrix}_{n_Z \times n_Z} \tag{3.19}
$$

where $n_Z = 2n_x + n_x(n_x + 1)/2$ is the dimension of Z_t, and M is an appropriately defined zero-one selection matrix of the form

$$
M = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m^* \end{bmatrix}
$$

The location of zeros and ones depends on the idiosyncratic microfoundations of the model at hand. For intuition, M is roughly a matrix that selects the states which have some persistence. The construction of the M matrix, featuring a simple

example, is discussed in Appendix C. The elements of the state equation A and B are written

$$A(\theta) = MP(\theta)M' \qquad B(\theta) = MR(\theta)N'$$

where R is the ancillary parameter

$$\underset{n_Z \times n_U}{R(\theta)} = \begin{bmatrix} \sigma\eta & 0 & 0 \\ 0 & \frac{1}{2}\sigma^2 H_{uu}D_{n_u} & \frac{1}{2}\sigma(H_{xu} + H_{ux}K_{n_x,n_u}) \\ 0 & \sigma^2 D_{n_x}^+ \eta^{\otimes 2}D_{n_u} & \sigma D_{n_x}^+(\eta \otimes h_x + (h_x \otimes \eta)K_{n_x,n_u}) \end{bmatrix} \tag{3.20}$$

for K_{n_x,n_u} the $n_x \times n_u$-dimensional square commutation matrix, and N another zero-one selection matrix of the form

$$N = \begin{bmatrix} I_{n_u} & 0 & 0 \\ 0 & I_{n_u} & 0 \\ 0 & 0 & n \end{bmatrix}$$

N also defines the error, by

$$\varepsilon_t = N \times \begin{bmatrix} u_t \\ D_{n_u}^+\left(u_t^{\otimes 2} - \text{vec}(I_{n_u^2})\right) \\ \text{vec}(\widehat{x}_{t-1}^f u_t') \end{bmatrix}$$

For intuition, N selects only the products within $\text{vec}(\widehat{x}_{t-1}^f u_t')$ for which the state element of the given product has persistence. Again, a full explanation of of how N is constructed is given in Appendix C. The variance-covariance matrix $\Sigma_\varepsilon(\theta) = E(\varepsilon_t\varepsilon_t'|\theta)$ is computed in Appendix A. Finally, the observables are defined as

$$Y_t = y_t^f + y_t^s - \left(\frac{1}{2}g_{\sigma\sigma}\sigma^2 + S(\theta) \times E(Z_t|\theta)\right)$$

for $S(\theta)$ the final necessary ancillary parameter

$$S(\theta) = \begin{bmatrix} g_x & g_x & \frac{1}{2}G_{xx}D_{n_x} \end{bmatrix}$$

The empirical analogue of Y_t is data that has been separated from means. Finally,

the matrices defining the observation equation are given in terms of the matrices above by

$$C(\theta) = S(\theta)P(\theta)M' \qquad D(\theta) = S(\theta)R(\theta)N'$$

While the transformation from pruned state space to ABCD evidently requires some rearranging, there are two immediately properties that are useful.

Corollary 1. *ε_t is $WN(0, \Sigma_\varepsilon(\theta))$ for finite $\Sigma_\varepsilon(\theta)$.*

Corollary 2. *The eigenvalues of A are less than one.*

Corollaries 1 and 2 are proven in Appendix A. These results emphasize that not only is it possible to rearrange the pruned state space to a representation which looks like something familiar from analysis of linearized models, two common assumptions for linearized DSGE model are also satisfied. There is one property of the above ABCD representation, however, that is less appealing.

Corollary 3. *$\{A, C\}$ is not observable.*

Corollary 3 is also proven in Appendix A. Given the ABCD representation of the pruned state space is non-observable, the most serious implication is that it is also therefore nonminimal.[5] Minimality is a key assumption of, for example, Komunjer and Ng (2011)'s rank and order conditions for identification in linearized models. Without it, such results are not applicable.

Although Komunjer and Ng's rank conditions pertaining to ABCD representation require minimality, the are also divided into two subsets of conditions: Minimal ABCD representation for singular models ($n_Y \leq n_\varepsilon$), and minimal AKCΣ representation for nonsingular models ($n_Y \geq n_\varepsilon$). Thus, the careful reader might initially guess that nonminimality of ABCD is not necessarily problematic, since models that have stochastically singular ABCD representation for linear approximations may have nonsingular ABCD representation of their corresponding pruned

---

[5]Definitions of key terms, such as *observability*, and its connection to *minimality*, are given Appendix D. These will be used throughout the rest of the paper. The concept of minimality comes from systems theory, and the definition in the Appendix is duplicated from a major textbook in that field, Kailath et al. (2000) page 765. See also Komunjer and Ng (2011) Definitions 5-S and 5-NS.

second order ABCD representation.[6] However, uncontrollability of ABCD also implies uncontrollability, and hence nonminimality, of AKCΣ. This may be seen by forming companion innovations form using one additional assumption.

Assumption 4. *For every $\theta \in \Theta$, $D(\theta)\Sigma_\varepsilon(\theta)D(\theta)'$ is nonsingular.*

Proposition 2. *Under Assumptions 1-4, the pruned state space (3.17) and (3.14) may written in terms of deviations from mean in $AKC\Sigma$ innovations representation*

$$
\begin{aligned}
\widehat{X}_{t|t} &= A(\theta) \ \widehat{X}_{t-1|t-1} \ + \ K(\theta)a_t \\
Y_t &= C(\theta) \ \widehat{X}_{t-1|t-1} \ + \ a_t
\end{aligned}
\tag{3.21}
$$

where $\widehat{X}_{t|t}$ is the optimal linear predictor of X_t given the history of observations, $a_t = Y_t - C\widehat{X}_{t|t}$ is the forecast error, and $K(\theta)$ is the steady state Kalman gain defined by

$$
K = \left(A\Sigma_X C' + B\Sigma_\varepsilon D'\right)\Sigma^{-1}
$$

where $\Sigma(\theta)$ is the covariance matrix of the forecast error

$$
\Sigma = C\Sigma_X C' + D\Sigma_\varepsilon D'
$$

and $\Sigma_X(\theta)$ is the covariance matrix of the state variables X_t defined by

$$
\Sigma_X = A\Sigma_X A' + B\Sigma_\varepsilon B' - \left(A\Sigma_X C' + B\Sigma_\varepsilon D'\right) \times \Sigma^{-1} \times \left(C\Sigma_X A' + D\Sigma_\varepsilon B'\right)
$$

This expression is known as the discrete algebraic Ricatti equation (DARE).

I provide only a short proof to Proposition 2 in Appendix A, since it follows directly from Proposition 1 when additionally Assumption 4 is satisfied, using well-known results. Further details on this closely related representation are available in Hansen and Sargent (2005).

---

[6]Explicitly, a linearized model is singular if $n_Y \geq n_u$. However, for pruned nonlinear models, the requirement for singularity is $n_Y \geq n_\varepsilon = (1 + n_m + (n_u + 1)/2)n_u$ where n_m is the row dimension of m in Appendix C. In most cases, $n_Y < n_\varepsilon$ even when $n_Y \geq n_u$. For example, consider the case where $n_Y = 6$ and $n_u = 2$. In linear ABCD representation, this model is easily singular. Now, assume $n_m = 1$. Then, $n_\varepsilon = (1 + n_m + (n_u + 1)/2)n_u = 7$, implying the pruned second order ABCD representation is not singular.

Clearly, because $\{A, C\}$ is known to not be observable, AKCΣ representation is also not minimal. Therefore, the implication of Proposition 2 is that the rank and order conditions derived in Komunjer and Ng are not immediately applicable. In the following section, I provide simple steps to obtain minimal ABCD and AKCΣ models starting from either nonminimal representation. Then, I provide a concrete example.

3.2 Constructing Minimal Representation

The reason why either ABCD or AKCΣ is not minimal is that it is not observable. Specifically, the currently defined observability matrix $\mathcal{O}$ will never be full column rank due to linear dependence between the first to block-columns of C; see, for example, the proof to Corollary 3 in Appendix A. It turns out that this problem is easy to amend. Define

$$\widehat{x}_t^{f+s} = \widehat{x}_t^f + \widehat{x}_t^s$$

It follows from the first two block-rows of the ABCD representation, given following Proposition 1, that

$$m\widehat{x}_t^{f+s} = \left(mh_x m'\right) m\widehat{x}_{t-1}^{f+s} + \left(\frac{1}{2}mH_{xx}D_{n_x}m^{*\prime}\right) m^* D_{n_x}^+ \widehat{x}_{t-1}^{f\otimes 2} + \sigma m\eta u_t$$

Therefore, the original ABCD system may immediately be rewritten as a new smaller dimensional ABCD model

$$
\underbrace{\begin{bmatrix} m\widehat{x}_t^{f+s} \\ m^* D_{n_x}^+ \widehat{x}_t^{f\otimes 2} \end{bmatrix}}_{X_t} = A(\theta) \begin{bmatrix} m\widehat{x}_{t-1}^{f+s} \\ m^* D_{n_x}^+ \widehat{x}_{t-1}^{f\otimes 2} \end{bmatrix} + B(\theta) \underbrace{\begin{bmatrix} u_t \\ D_{n_u}^+ \left(u_t^{\otimes 2} - \text{vec}(I_{n_u^2})\right) \\ n\text{vec}(\widehat{x}_{t-1}^f u_t') \end{bmatrix}}_{\varepsilon_t}
$$

$$
Y_t = C(\theta) \begin{bmatrix} m\widehat{x}_{t-1}^{f+s} \\ m^* D_{n_x}^+ \widehat{x}_{t-1}^{f\otimes 2} \end{bmatrix} + D(\theta) \begin{bmatrix} u_t \\ D_{n_u}^+ \left(u_t^{\otimes 2} - \text{vec}(I_{n_u^2})\right) \\ n\text{vec}(\widehat{x}_{t-1}^f u_t') \end{bmatrix}
$$

$$(3.22)$$

where

$$A = \left[\begin{array}{c|c} mh_x m' & \frac{1}{2} m H_{xx} D_{n_x} m^{*'} \\ \hline 0 & m D_{n_x}^+ h_x^{\otimes 2} D_{n_x} m^{*'} \end{array} \right]$$

$$B = \left[\begin{array}{c|c|c} \sigma m \eta & \frac{1}{2} \sigma^2 m H_{uu} D_{n_u} & \frac{1}{2} \sigma (H_{xu} + H_{ux} K_{n_x,n_u}) n' \\ \hline 0 & \sigma^2 m D_{n_x}^+ \eta^{\otimes 2} D_{n_u} & \sigma m D_{n_x}^+ (\eta \otimes h_x + (h_x \otimes \eta) K_{n_x,n_u}) n' \end{array} \right]$$

$$C(\theta) = \left[\begin{array}{c|c} g_x h_x m' & \frac{1}{2} \left(g_x H_{xx} + G_{xx} h_x^{\otimes 2} \right) D_{n_x} m^{*'} \end{array} \right]$$

$$D(\theta) = \left[\begin{array}{c|c|c} \sigma g_x \eta & \frac{1}{2} \sigma^2 \left(g_x H_{uu} D_{n_u} + G_{xx} \eta^{\otimes 2} D_{n_u} \right) & d_3 \end{array} \right]$$

and

$$d_3 = \frac{1}{2} \sigma \left(g_x (H_{xu} + H_{ux} K_{n_x,n_u}) + G_{xx} (\eta \otimes h_x + (h_x \otimes \eta) K_{n_x,n_u}) \right) n'$$

In fact, in many cases, this new ABCD system (and corresponding AKCΣ) is observable, and minimal. In order to substantiate this claim, and to clarify the derivations thus far, in the next section I provide a simple example.

3.3   Example: Schmitt-Grohé and Uribe (2004)

To make the methodology for minimal reparameterization of pruned nonlinear models concrete, in this section I show to obtain ABCD and AKCΣ representation using a microfounded example. Consider the simple neoclassical growth model studied in Schmitt-Grohé and Uribe (2004):

$$c_t^{-\gamma} = \beta E_t \{ c_{t+1}^{-\gamma} \left(\alpha A_{t+1} k_{t+1}^{\alpha-1} + 1 - \delta \right) \} \tag{3.23}$$

$$c_t + k_{t+1} = A_t k_t^{\alpha} + (1 - \delta) k_t \tag{3.24}$$

$$\ln A_{t+1} = \rho \ln A_t + \sigma u_{t+1} \tag{3.25}$$

k_t is capital, A_t is total factor productivity, and c_t is the only observable variable, consumption. For the purposes of this example, I will assume the scalar shock u_{t+1} is standard normal. The structural parameters of the model are collected in the 5×1 vector $\theta = (\beta, \delta, \alpha, \rho, \gamma)'$, and σ is the perturbation parameter. Due

to nonlinearity between states and errors in the Euler equation, the states are defined to be $x_t = \begin{bmatrix} k_t & c_t \end{bmatrix}'$ the single observable is $y_t = c_t$, and the single shock is $u_t \sim WN(0,1)$. The parameter value studied in Schmitt-Grohé and Uribe is

$$\underset{5 \times 1}{\theta_0} = \Big(\ \beta = 0.95\ \vdots\ \delta = 1\ \vdots\ \alpha = 0.3\ \vdots\ \rho = 0\ \vdots\ \gamma = 2\ \Big)' \tag{3.26}$$

Let us consider then the 5-dimensional parameter space $\theta \in \Theta$ in which $\rho = 0$ always, and all other parameters are both real-valued, and imply a solution exists. Given this restriction, the model is written in the form of (3.15) as

$$\begin{bmatrix} k_t \\ c_t \\ \ln A_{t+1} \end{bmatrix} = \begin{bmatrix} \underset{2 \times 1}{h(k_{t-1}, c_{t-1}, \ln A_t | \theta)} \\ \hline 0 \end{bmatrix} + \sigma \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u_{t+1}$$

Note, $\ln A_{t+1}$ is playing the role of v_{t+1} in Equation (3.15). The (rounded) values of the solution in the form of equation (3.16) are[7]

$$\underbrace{\begin{bmatrix} \widehat{k}_t \\ \widehat{c}_t \end{bmatrix}}_{\widehat{x}_t} \approx \underbrace{\begin{bmatrix} 0.42 & 0 \\ 0.25 & 0 \end{bmatrix}}_{h_x(\theta_0)} \underbrace{\begin{bmatrix} \widehat{k}_{t-1} \\ \widehat{c}_{t-1} \end{bmatrix}}_{\widehat{x}_{t-1}} + \underbrace{\begin{bmatrix} -0.004 & 0 & 0 & 0 \\ -0.003 & 0 & 0 & 0 \end{bmatrix}}_{\frac{1}{2}H_{xx}(\theta_0)} \underbrace{\begin{bmatrix} \widehat{k}_{t-1}^2 \\ \widehat{c}_{t-1}\widehat{k}_{t-1} \\ \widehat{k}_{t-1}\widehat{c}_{t-1} \\ \widehat{c}_{t-1}^2 \end{bmatrix}}_{\widehat{x}_{t-1}^{\otimes 2}}$$

$$+ \underbrace{\begin{bmatrix} -0.012 & 0 \\ -0.008 & 0 \end{bmatrix}}_{\frac{1}{2}\sigma H_{xu}(\theta_0)} \underbrace{\begin{bmatrix} \widehat{k}_{t-1}u_t \\ \widehat{c}_{t-1}u_t \end{bmatrix}}_{\text{vec}(\widehat{x}_{t-1}u_t')} + \underbrace{\begin{bmatrix} -0.04 \\ -0.03 \end{bmatrix}}_{\frac{1}{2}\sigma^2 H_{uu}(\theta_0)} u_t^2 + \underbrace{\begin{bmatrix} -0.012 & 0 \\ -0.008 & 0 \end{bmatrix}}_{\frac{1}{2}\sigma H_{ux}(\theta_0)} \underbrace{\begin{bmatrix} \widehat{k}_{t-1}u_t \\ \widehat{c}_{t-1}u_t \end{bmatrix}}_{\text{vec}(u_t\widehat{x}_{t-1}')}$$

$$+ \underbrace{\begin{bmatrix} 0.24 \\ -0.1 \end{bmatrix}}_{\frac{1}{2}h_{\sigma\sigma}(\theta_0)\sigma^2} + \underbrace{\begin{bmatrix} 1.39 \\ 0.84 \end{bmatrix}}_{\sigma\eta(\theta_0)} u_t \tag{3.27}$$

---

[7]These are computed using the Dynare code SGU2004.mod provided freely online by Johannes Pfeifer.

and the observation equation is merely an identity.

$$\underbrace{\widehat{c_t}}_{\widehat{y}_t} = \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix}}_{g_x(\theta_0)} \widehat{x}_t + \underbrace{0}_{\frac{1}{2}G_{xx}(\theta_0)} \widehat{x}_t^{\otimes 2} + \underbrace{0}_{\frac{1}{2}g_{\sigma\sigma}(\theta)\sigma^2} \tag{3.28}$$

The zeros in h_x and H_{xx} arise due to the fact that TFP is not persistent, $\rho = 0$. Since we are considering the parameter space Θ in which $\rho = 0$ always, Assumption 2 is satisfied. In order to simplify the model at hand to ABCD representation, I now use the three-step methodology described in Appendix C.

Step 1. The pruned model may be represented by the rule of motion

$$Z_t = J(\theta_0) + P(\theta_0)Z_{t-1} + R(\theta_0)U_t \tag{3.29}$$

$$\mathcal{Y}_t = S(\theta_0)Z_t \tag{3.30}$$

where J, P, and R correspond to the expressions given previously in equations (3.19) and (3.20).

$$\underset{7\times 1}{Z_t} = \begin{bmatrix} k_t^f & c_t^f & k_t^s & c_t^s & (k_t^f)^2 & c_t^f k_t^f & (c_t^f)^2 \end{bmatrix}'$$

$$\underset{1\times 1}{\mathcal{Y}_t} = c_t^f + c_t^s \qquad \underset{4\times 1}{U_t} = \begin{bmatrix} u_t & u_t^2 - 1 & \widehat{k}_{t-1}u_t & \widehat{c}_{t-1}u_t \end{bmatrix}'$$

$$\underset{7\times 1}{J(\theta_0)} = \begin{bmatrix} 0 & 0 & 0.20 & -0.13 & 1.95 & 1.18 & 0.71 \end{bmatrix}'$$

$$\underset{7\times 7}{P(\theta_0)} = \begin{bmatrix} 0.42 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.42 & 0 & -0.004 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & -0.003 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.18 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.11 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.06 & 0 & 0 \end{bmatrix} \qquad \underset{7\times 4}{R(\theta_0)} = \begin{bmatrix} 1.40 & 0 & 0 & 0 \\ 0 & -0.04 & -0.02 & 0 \\ 0 & -0.03 & -0.02 & 0 \\ 0 & 1.95 & 1.17 & 0 \\ 0 & 1.18 & 0.71 & 0 \\ 0 & 0.71 & 0.43 & 0 \end{bmatrix}$$

$$S(\theta_0) = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Step 2. The expected value of Z_t is defined by

$$\underset{7\times 1}{E(Z_t|\theta_0)} = (I_7 - P(\theta_0))^{-1} J(\theta_0) = \begin{bmatrix} 0 & 0 & 0.35 & -0.04 & 2.37 & 1.43 & 0.86 \end{bmatrix}'$$

Therefore, equations (3.29) and (3.30) may be written

$$\widehat{Z}_t = P(\theta_0)\widehat{Z}_{t-1} + R(\theta_0)U_t \tag{3.31}$$

$$Y_t = S(\theta_0)\widehat{Z}_t \tag{3.32}$$

for $\widehat{Z}_t = Z_t - E(Z_t|\theta_0)$ and $Y_t = \mathcal{Y}_t - S(\theta_0)E(Z_t|\theta_0)$.

Step 3. Using the positions of zeros in P and R as guidelines, define

$$m = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad m^* = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \quad \underset{3\times 7}{M} = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m^* \end{bmatrix}$$

$$n = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad \underset{3\times 4}{N} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & n \end{bmatrix}$$

Using M and N as defined above, we have ABCD representation, equation (3.18) with

$$\underset{3\times 1}{X_t} = M \times \widehat{Z}_t = \begin{bmatrix} \widehat{k}_t^f \\ \widehat{k}_t^s - 0.35 \\ (\widehat{k}_t^f)^2 - 2.37 \end{bmatrix} \quad \underset{3\times 1}{\varepsilon_t} = N \times U_t = \begin{bmatrix} u_t \\ u_t^2 - 1 \\ \widehat{k}_{t-1}^f u_t \end{bmatrix}$$

$$\underset{3\times 3}{A(\theta_0)} = \begin{bmatrix} 0.42 & 0 & 0 \\ 0 & 0.42 & -0.004 \\ 0 & 0 & 0.18 \end{bmatrix} \quad \underset{3\times 4}{B(\theta_0)} = \begin{bmatrix} 1.40 & 0 & 0 \\ 0 & -0.04 & -0.02 \\ 0 & 1.95 & 1.17 \end{bmatrix}$$
$$\qquad\qquad MP(\theta_0)M' \qquad\qquad\qquad\qquad MR(\theta_0)N'$$

$$\underset{1\times 3}{C(\theta_0)} = \begin{bmatrix} 0.25 & 0.25 & -0.003 \end{bmatrix} \quad \underset{1\times 3}{D(\theta_0)} = \begin{bmatrix} 0.84 & -0.03 & -0.02 \end{bmatrix}$$
$$\qquad\quad SP(\theta_0)M' \qquad\qquad\qquad\quad SR(\theta_0)N'$$

Assume $E(u_t^3) = 0$ and $E(u_t^4 = 3)$; this is true, for example, if u_t is standard

normal. In addition, observe that the rule of motion for $\widehat{k}_t^f$ evaluated at θ_0 is $\widehat{k}_t^f = 0.42\widehat{k}_{t-1}^f + 1.4u_t$, we have $E(\widehat{k}_t^{f2}|\theta_0) = 1.4^2/(1 - 0.42^2) = 2.37$. Then the covariance matrix of the errors is written

$$\underset{3\times3}{\Sigma_\varepsilon(\theta_0)} = E(\varepsilon_t\varepsilon_t'|\theta_0) = \begin{bmatrix} 1 & \cdot & \cdot \\ 0 & 2 & \cdot \\ 0 & 0 & 2.37 \end{bmatrix}$$

Minimal Representation. The controllability matrix $\mathcal{C}(\theta_0)$ for the ABCD system is full row rank, implying it is controllable. However, the observability matrix is

$$\underset{3\times3}{\mathcal{O}(\theta_0)} = \begin{bmatrix} C \\ C \times A \\ C \times A^2 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.25 & -0.003 \\ 0.11 & 0.11 & -0.001 \\ 0.04 & 0.04 & -0.001 \end{bmatrix}$$

A necessary and sufficient requirement for observability, and hence minimality, is that $\mathcal{O}$ is full column rank. However, it is clear that the first two columns are linearly dependent. Using the suggested approach from Section 3.2 gives

$$\underbrace{\begin{bmatrix} \widehat{k}_t^f + \widehat{k}_t^s - 0.35 \\ (\widehat{k}_t^f)^2 - 2.37 \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} 0.42 & -0.004 \\ 0 & 0.18 \end{bmatrix}}_{A(\theta)} X_{t-1} + \underbrace{\begin{bmatrix} 1.40 & -0.04 & -0.02 \\ 0 & 1.95 & 1.17 \end{bmatrix}}_{B(\theta_0)} \varepsilon_t$$

$$Y_t = \underbrace{\begin{bmatrix} 0.25 & -0.003 \end{bmatrix}}_{C(\theta_0)} X_{t-1} + \underbrace{\begin{bmatrix} 0.84 & -0.02 & -0.02 \end{bmatrix}}_{D(\theta_0)} \varepsilon_t$$

$$(3.33)$$

This ABCD system is both controllable and observable, and hence minimal. Finally, AKCΣ innovations representation is easily constructed by solving the DARE, yielding $\Sigma(\theta_0) = 0.71$ and $K(\theta_0) = \begin{bmatrix} 1.66 & -0.23 \end{bmatrix}'$. This representation is also minimal.

3.4 A Key Property: Nestedness

The first order minimal ABCD representation of the SGU model is in fact nested within minimal nonlinear representation, Equation (3.33). It is given by

$$
\begin{aligned}
\widehat{k}_t^f &= 0.42\widehat{k}_{t-1}^f &+& 1.40u_t \\
\widehat{c}_t^f &= 0.25\widehat{k}_{t-1}^f &+& 0.84u_t
\end{aligned}
\tag{3.34}
$$

with the linear version of innovations representation $K(\theta_0) = 1.66$ and $\Sigma(\theta_0) = 0.71$ also nested within the nonlinear versions of $K(\theta_0)$ and $\Sigma(\theta_0)$. Ultimately, this is due to the fact that higher-order Taylor approximations nest lower-order approximations. Since none of the matrices A, B, C, D, Σ_ε, K, or Σ are themselves identifiable, it is impossible to conclude that higher order models will always procure the identification of weakly more parameters than nonlinear. But clearly, this outcome is worth consideration.

In the next section, I consider the local identification of nonlinear models versus their linear counterparts, using the minimal representation of nonlinear models developed in this section, and repurposing the rank and order conditions derived for minimal linearized models by Komunjer and Ng (2011) to this case. I first show that exactly because of the nested nature of the higher order Schmitt-Grohé and Uribe solution, the identification of strictly more parameters becomes possible. Finally, I conduct a full analysis of the model of An and Schorfheide (2007), which was also studied in Komunjer and Ng. Again, I observe a nested nature of the model.

4 Local Identification

Komunjer and Ng (2011) derive necessary order conditions, and necessary and sufficient rank conditions, for local identification of the structural parameters in DSGE models. These conditions are based on identifiability from the spectral density, i.e. the sequence of first and second moments of an infinitely long data set. The results are split into two sets of conditions, one for stochastically singular models $n_Y \geq n_\varepsilon$, and the second for stochastically nonsingular models $n_Y \leq n_\varepsilon$; the

two results coincide only if $n_Y = n_\varepsilon$. The conditions framed around the singular case are based on minimality of ABCD representation, while the conditions under nonsingularity rely on minimal AKCΣ representation. In either case, the results are intended for linearized DSGE models, which are easy to represent in such functional forms.

In the previous section, I have shown that pruned second order approximations of DSGE model solutions also may be expressed in minimal ABCD or AKCΣ representation. In these models, the dimension of the error is $n_\varepsilon = (1 + n_m)n_u + n_u(n_u + 1)/2$, where n_m is the dimension of the matrix m. Given this fact, in most cases, $n_\varepsilon > n_Y$. Therefore, I will direct attention to this set of rank and order conditions directed towards minimal AKCΣ representation.

In order to make this set of KN conditions operable, we require one final assumption.

Assumption 5. *The mapping* $\Lambda : \theta \to \Lambda(\theta)$ *is continuously differentiable on* Θ *where*

$$\Lambda(\theta) = \left[\; vec\left(A(\theta)\right)' \; \vdots \; vec\left(K(\theta)\right)' \; \vdots \; vec\left(C(\theta)\right)' \; \vdots \; vech\left(\Sigma(\theta)\right)' \; \right]'$$

Furthermore, define the following matrix of derivatives end elements of Λ:

$$\Delta(\theta) = \left[\begin{array}{c|c} \frac{\partial \mathrm{vec} A(\theta)}{\partial \theta'} & A(\theta)' \otimes I_{n_X} - I_{n_X} \otimes A(\theta) \\ \hline \frac{\partial \mathrm{vec} K(\theta)}{\partial \theta'} & K(\theta)' \otimes I_{n_X} \\ \hline \frac{\partial \mathrm{vec} C(\theta)}{\partial \theta'} & -I_{n_X} \otimes C(\theta) \\ \hline \frac{\partial \mathrm{vech} \Sigma(\theta)}{\partial \theta'} & 0_{n_Y(n_Y+1)/2 \times n_X^2} \end{array} \right] \tag{3.35}$$

Then, we have the following immediate result, a verbatim statement of Komunjer and Ng Proposition 2-NS. Recall, under Assumptions 1-3, Corollary 1 ensures ε_t is white noise.

Proposition 3. *Suppose* $n_\varepsilon \geq n_Y$ *and Assumptions 1-5 hold. If the rank of* $\Delta(\theta)$ *remains constant in a neighborhood of* θ_0*, then a necessary and sufficient condition*

for θ to be locally identified from the autocovariances of Y_t at a point θ_0 in Θ is rank $\Delta(\theta_0) = n_\theta + n_X$. A necessary order condition is $n_\theta \leq 2n_X n_Y + n_Y(n_Y + 1)/2$.

Because the leg-work of reparameterizing the nonlinear model was accomplished in the previous section, in this section, we are able to concentrate on the implications of Proposition 3 for the identifiability of nonlinear models. In particular, one item of interest is how the identifiability of parameters in a second order pruned approximation of a model compares with the identifiability of parameters in a linear approximation of the same model. As we have seen, a relic of the nestedness of Taylor approximations is that minimal ABCD and AKCΣ representation of the linearized Schmitt-Grohé and Uribe model (3.34) is contained within the ABCD and AKCΣ representation of the second order pruned model (3.33). Thus, intuition suggests that identifiability of certain parameters may be enhanced in the higher order case. Yet, the validity of this hypothesis must be tested.

In the next subsection, I study the identifiability of the parameters of the Schmitt-Grohé and Uribe model using a linear versus nonlinear approximation. I confirm the hypothesis that strictly more parameters are identifiable in the nonlinear model. In the following section, I turn attention to the model of An and Schorfheide (2007), the linear version of which is also studied by Komunjer and Ng. Beginning with the identifying parameter restrictions Komunjer and Ng suggest, I show that nonlinearity allows one to relax a subset of these restrictions. I confirm An and Schorfheide's intuition that three important macroeconomic parameters become identifiable in a nonlinear approximation which were otherwise not identifiable using a linear approximation.

4.1 Example: Schmitt-Grohé and Uribe (2004)

As I have shown, the linearized model of the Schmitt-Grohé and Uribe model is nested within the second order pruned model. In Figure 3.1 I provide values for both Λ and Δ evaluated at θ_0 for both the linear model (3.34) and nonlinear model (3.33).

First, consider the statistics corresponding to the linear approximation. In

First-order approximation. Order condition not satisfied. $\text{Rank}(\Delta(\theta_0)) = 4$.

$$\Lambda(\theta_0) = \begin{bmatrix} 0.42 & 1.66 & 0.25 & 0.71 \end{bmatrix}'$$
$$\scriptstyle 4\times 1$$

$$\underset{4\times 6}{\Delta(\theta_0)} = \begin{bmatrix} 0.09 & -0.33 & 1.07 & -0.05 & 0 & 0 \\ 0.55 & -0.56 & -0.70 & 0.13 & -0.65 & 1.65 \\ -0.04 & -0.10 & 0.75 & -0.05 & 0 & -0.25 \\ -0.18 & 0.71 & -0.50 & -0.28 & 0.22 & 0 \end{bmatrix}$$

Second-order pruned approximation. $\text{Rank}(\Delta(\theta_0)) = 9$, achieving rank condition. Above linearized model statistics are nested in **bold**.

$$\underset{9\times 1}{\Lambda(\theta_0)} = \begin{bmatrix} \mathbf{0.42} & 0 & -0.004 & 0.18 & \mathbf{1.66} & -0.23 & \mathbf{0.25} & -0.003 & \mathbf{0.71} \end{bmatrix}'$$

$$\underset{9\times 9}{\Delta(\theta_0)} = \begin{bmatrix} \mathbf{0.09} & \mathbf{-0.33} & \mathbf{1.07} & \mathbf{-0.05} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.24 & 0 & 0 \\ -0.01 & -0.10 & -0.02 & 0 & 0 & 0 & 0 & -0.24 & 0 \\ 0.07 & -0.28 & 0.90 & -0.04 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{0.55} & \mathbf{-0.56} & \mathbf{-0.70} & \mathbf{0.13} & \mathbf{-0.65} & \mathbf{1.65} & 0 & 0.13 & 0 \\ 0.27 & -0.94 & -0.16 & 0.08 & -0.25 & 0 & 1.65 & 0 & 0.13 \\ \mathbf{-0.04} & \mathbf{-0.10} & \mathbf{0.75} & \mathbf{-0.05} & \mathbf{0} & \mathbf{-0.25} & 0 & 0 & 0 \\ 0 & -0.05 & -0.02 & 0 & 0 & 0 & 0 & -0.25 & 0 \\ \mathbf{-0.18} & \mathbf{0.71} & \mathbf{-0.50} & \mathbf{-0.28} & \mathbf{0.22} & \mathbf{0} & 0 & 0 & 0 \end{bmatrix}$$

Figure 3.1: Identification of Schmitt-Grohé and Uribe model: Linear versus second order pruned approximation.

this model, $n_X = n_Y = 1$ and $n_\theta = 5$. So, $2n_X n_Y + n_Y(n_Y + 1)/2 = 3 < n_\theta = 5$ implies the necessary order condition in Proposition 3 is not satisfied. Thus, without even considering the rank of Δ, it follows that the entire 5-dimensional vector θ is not identifiable, and at least two parameters must be set to constants for their complement in θ to be conditionally identified.

Now, consider the same statistics Λ and Δ for the second order pruned version of the SGU model, also given in Figure 3.1. The elements of the 4×1 dimensional Λ from the first order approximation are contained within the 9×1 dimensional Λ for this version, in bold. Since $n_X = 2$ and $n_Y=1$, we have $2n_X n_Y + n_Y(n_Y + 1)/2 = 5 = n_\theta$, thus satisfying the necessary order condition for identifiability, unlike the linearized model. More interesting, though, is the rela-

tionship between Δ for the linear and nonlinear approximations; again, I highlight the feature of nestedness using bold. In fact, in the second order pruned version of the Schmitt-Grohé and Uribe model, the 9×9 matrix $\Delta(\theta)$ is full column rank at θ_0, satisfying the necessary and sufficient condition for local identification of the entire vector θ at θ_0 in Proposition 3.

The SGU model has helped illustrate the important insight of this paper: Higher order approximations of DSGE models nest lower-order approximations, and the additional terms from nonlinear approximations can help identify key macroeconomic parameters. In the case of the Schmitt-Grohé and Uribe model, at most 3 parameters are identifiable using a linear approximation, but all 5 are identified at θ_0 by simply using a nonlinear approximation.

As I have shown, analysts might be able to overcome identification problems faced in linearized models simply by using higher-order approximations. While general statements to this effect are difficult to make, the improved identifiability of DSGE models from nonlinear approximations is not limited to the simple Schmitt-Grohé and Uribe model. In fact, the pruned state space approximation can also be used to identify more parameters in more empirically plausible specifications, like An and Schorfheide (2007)'s. I demonstrate this fact in the following section.

5 Application: An and Schorfheide (2007)

The An and Schorfheide model, which includes the baseline elements of many DSGE models, is well-known and well-studied. The Appendix to the first chapter of this dissertation derives the nonlinear equilibrium equations of the model of this model as

$$
\begin{aligned}
\frac{1}{\nu}\left(1 - \exp\left\{\tau \ln \widehat{C}_t\right\}\right) &+ \phi(\exp\{\ln \Pi_t\} - \Pi)\left(\left(1 - \frac{1}{2\nu}\right)\exp\{\ln \Pi_t\} + \frac{\Pi}{2\nu}\right) - \\
\phi\beta E_t &\left[\exp\left\{\ln(\widehat{Y}_{t+1}) - \ln(\widehat{Y}_t) - \tau\left(\ln(\widehat{C}_{t+1}) - \ln(\widehat{C}_t)\right)\right\} \times \right.\\
&\left. (\exp\{\ln \Pi_{t+1}\} - \Pi)\exp\{\ln \Pi_t\}\right] - 1 = 0 \quad (3.36)
\end{aligned}
$$

Table 3.1: An and Schorfheide (2007) model parameter and variable names.

		Structural Params (15)	Endogenous (6)		Shocks (3)	
1	τ	CRRA	Z_t	TFP	ε_{zt}	TFP
2	β	Discount factor	G_t	Gov spending	ε_{gt}	Gov.
3	ν	Inverse elas. of demnd	R_t	Nominal int rate	ε_{rt}	Int.
4	ϕ	Index of price stckness	$\widehat{Y}_t$	Nom. detr. output		
5	γ	Avg. gr. rate of prod.	Π_t	Inflation		
6	Π	St. state level of infl.	$\widehat{C}_t$	Nom. detr. 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				

$$\beta E_t \left[\exp \left\{ \ln R_t - \ln \Pi_{t+1} - \tau \left(\ln(\widehat{C}_{t+1}) - \ln(\widehat{C}_t) \right) - \ln \gamma - \ln Z_{t+1} \right\} \right] - 1 = 0 \tag{3.37}$$

$$- \ln R_t + (1 - \rho_r) \left(\ln \gamma - \ln \beta + (1 - \psi_\pi) \ln \Pi - \psi_y (1/\tau) \ln(1 - \nu) \right) + \rho_r \ln R_{t-1}$$
$$+ (1 - \rho_r) \psi_\pi \ln \Pi_t + (1 - \rho_r) \psi_y (\ln \widehat{Y}_t - \ln G_t) + \sigma_r \varepsilon_{rt} = 0 \tag{3.38}$$

$$- \exp \left\{ \ln \widehat{Y}_t \right\} + \exp \left\{ \ln \widehat{C}_t + \ln G_t \right\} \bigg/ \left(1 - \frac{\phi}{2} (\exp \{ \ln \Pi_t \} \right.$$
$$\left. - \Pi)^2 \exp \{ \ln G_t \} \right) = 0 \tag{3.39}$$

$$- \ln Z_t + \rho_z \ln Z_{t-1} + \sigma_z \varepsilon_{zt} = 0 \tag{3.40}$$

$$- \ln G_t + (1 - \rho_g) \ln G + \rho_g \ln G_{t-1} + \sigma_g \varepsilon_{gt} = 0 \tag{3.41}$$

All variables and structural parameters of this model are given in Table 3.1. The shock vector $u_t = [\varepsilon_{z_t}, \varepsilon_{gt}, \varepsilon_{rt}]'$ is distributed as $u_t \sim WN(0, I_3)$. In all, there are six key equilibrium equations (A.17) - (A.22) that completely 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}. Henceforth, these total nine variables will be collected in the following vectors, where v_t is the ancillary parameter for models with nonlinearities in errors and shocks defined in (3.15).

$$\underset{(6\times1)}{\left[x'_{t-1} \quad v'_t\right]'} = \left[\ln Z_{t-1} \quad \ln G_{t-1} \quad \ln R_{t-1} \quad \varepsilon_{zt} \quad \varepsilon_{gt} \quad \varepsilon_{rt}\right]' \tag{3.42}$$

$$\underset{(3\times1)}{y_t} = \left[\ln R_t \quad \ln \hat{Y}_t \quad \ln \Pi_t\right]' \tag{3.43}$$

With these definitions in mind, the equilibrium equations may also be represented concisely in the form of Equation (3.1). Thus, the solution will have the form of Equations (3.2) and (3.15). Define $h_{(3)}$ to be the third row of h. Then the state equation may be expressed as

$$\underbrace{\begin{bmatrix} \ln Z_t \\ \ln G_t \\ \ln R_t \\ \varepsilon_{zt+1} \\ \varepsilon_{gt+1} \\ \varepsilon_{rt+1} \end{bmatrix}}_{[x'_t \; v'_{t+1}]'} = \begin{bmatrix} \rho_z \ln Z_{t-1} + \varepsilon_{zt} \\ \rho_g \ln G_{t-1} + \varepsilon_{gt} \\ h_{(3)}(x_{t-1}, v_t, \sigma|\theta) \\ 0 \\ 0 \\ 0 \end{bmatrix} + \sigma \underbrace{\begin{bmatrix} 0_{3\times3} \\ \underbrace{\begin{bmatrix} \sigma_z & \cdot & \cdot \\ 0 & \sigma_g & \cdot \\ 0 & 0 & \sigma_r \end{bmatrix}}_{L_u(\theta)} \end{bmatrix}}_{\eta(\theta)} \underbrace{\begin{bmatrix} \varepsilon_{zt+1} \\ \varepsilon_{gt+1} \\ \varepsilon_{rt+1} \end{bmatrix}}_{u_{t+1}} \tag{3.44}$$

Definining $g_{(i)}$ to be the i-th row of g, the observation equation is simply stated as

$$\underbrace{\begin{bmatrix} \ln R_t \\ \ln \hat{Y}_t \\ \ln \Pi_t \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} \ln R_t \\ g_{(2)}(x_t, \sigma|\theta) \\ g_{(3)}(x_t, \sigma|\theta) \end{bmatrix}}_{g(x_t, \sigma|\theta)} \tag{3.45}$$

The value of the structural parameters at which I will study identification is θ_0, given in Table 3.2. Solving the model and using the same 3-step procedure given

Table 3.2: Candidate parameter point θ_0.

	Parameter	θ_0
1	τ	2
2	β	0.9975
3	ν	0.1
4	ϕ	53.68
5	γ	1.002
6	Π	1.008
7	G	1.18
8	ψ_π	1.5
9	ψ_y	0.125
10	ρ_z	0.9
11	ρ_g	0.95
12	ρ_r	0.75
13	σ_z	0.003
14	σ_g	0.006
15	σ_r	0.002

in Appendix C and demonstrated in the context of the Schmitt-Grohé and Uribe model in Section 3.3 yields the minimal ABCD representation in Table 3.2. z_t, g_t, and r_t are first-order approximations of TFP, government spending, and interest rates. Second-order terms are collected into the 5×1-dimensional $\widetilde{x}_t$ and 13×1-dimensional $\widetilde{\varepsilon}_t$.

Compare this ABCD representation with the minimal ABCD representation of the linearized version of the same model at θ_0 in Komunjer and Ng (2011) Table 1. Similarly to the Schmitt-Grohé and Uribe model, in this case, it is quite clear how the linear model nested within the nonlinear approximation. Therefore, a natural question to ask is whether more parameters are identifiable in this case as well. In Table 3.2 I also consider six sets of restrictions on the parameters of the model, and whether the complement set is identifiable in a linear, or nonlinear model. Restrictions 1 and 2, which are both restrictions for 5 parameters, correspond to those Komunjer and Ng show successfully identify the complement 13. These restrictions also work for the nonlinear model. As Komunjer and Ng show, this is the bare minimum amount of parameters which must be fixed for the complement set to be conditionally identified.

Although many parameters are not identifiable in the linearized model,

$$
\underbrace{\begin{bmatrix} z_t \\ g_t \\ r_t \\ \widetilde{x}_t \end{bmatrix}}_{\substack{X_t \\ 8\times 1}}
=
\underbrace{\begin{bmatrix} 0.9 & 0 & 0 & A_{14} \\ 0 & 0.95 & 0 & A_{24} \\ 0.55 & 0 & 0.51 & A_{34} \\ 0 & 0 & 0 & A_{44} \end{bmatrix}}_{\substack{A(\theta_0) \\ 8\times 8}}
\underbrace{\begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \\ \widetilde{x}_{t-1} \end{bmatrix}}_{X_t}
+
\underbrace{\begin{bmatrix} 1 & 0 & 0 & B_{14} \\ 0 & 1 & 0 & B_{24} \\ 0.61 & 0 & 0.69 & B_{34} \\ 0 & 0 & 0 & B_{44} \end{bmatrix}}_{\substack{B(\theta_0) \\ 8\times 16}}
\underbrace{\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \\ \widetilde{\varepsilon}_t \end{bmatrix}}_{\substack{\varepsilon_t \\ 18\times 1}}
$$

$$
\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t}
=
\underbrace{\begin{bmatrix} 0.55 & 0 & 0.51 & C_{14} \\ 1.34 & 0.95 & -0.83 & C_{24} \\ 1.34 & 0 & -0.56 & C_{34} \end{bmatrix}}_{\substack{C(\theta_0) \\ 3\times 8}}
\underbrace{\begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \\ \widetilde{x}_{t-1} \end{bmatrix}}_{X_t}
+
\underbrace{\begin{bmatrix} 0.61 & 0 & 0.69 & D_{14} \\ 1.49 & 1 & -1.1 & D_{24} \\ 1.49 & 0 & -0.75 & D_{34} \end{bmatrix}}_{\substack{D(\theta_0) \\ 3\times 16}}
\underbrace{\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{rt} \\ \widetilde{\varepsilon}_t \end{bmatrix}}_{\varepsilon_t}
$$

	Restriction					Linear	Nonlinear	Additional
1.	ν	ϕ	ψ_π	G	γ	✓	✓	—
2.	ν	ϕ	ψ_y	G	γ	✓	✓	—
3.	—	ϕ	ψ_π	ψ_y	γ		✓	ν
4.	—	ν	ψ_π	ψ_y	γ		✓	ϕ
5.	—	ν	ϕ	ψ_y	γ		✓	G
6.	—	ν	ϕ	ψ_y	G			—

Figure 3.2: Minimal ABCD representation of the second order pruned solution of the An and Schorfheide model.

An and Schorfheide claim in their paper that the likelihood profile for a nonlinear approximation of this model shows curvature in all of ν, ϕ, and G, whereas a linear model does not. They hypothesize that this means that those three parameters may be identifiable in a nonlinear version of the model, but not linear. Are they correct in this assertion?

Yes. In Table 3.2, I consider three more sets of restrictions, 3, 4, and 5. Restriction 3 allows ν to be free, and sets only 4 parameters to constants. In this case, the nonlinear model successfully identifies the additional parameter ν. Restriction 4 allows the identification of the additional parameter ϕ. Finally, Restriction 5 allows for the identification of the additional parameter G. Therefore, I am able to verify An and Schorfheide's intuition that all three of these parameters are identifiable in a nonlinear version of the model, but not linear. It is important to realize, however, that nonlinear models do not provide a silver bullet with respect to identification. For instance, Restriction 6 does not result in the identification of the complement set in θ.

6   Conclusion

In this paper, I have shown how to assess parameter identifiability in nonlinear approximations of DSGE models. Due to the inherent nestedness of Taylor approximations, nonlinear approximations of these models may be used to identify key parameters of interest that are otherwise not identifiable from a linear approximation of the same model. In the context of the An and Schorfheide (2007) model, I have shown this to be true for three important parameters, the elasticity of substitution, price stickiness, and steady state level of government spending. Yet, even in nonlinear models, a subset of parameters typically must be restricted to identify their complement. This paper has at the same time introduced a pragmatic methodology for determining which.

Appendix A

Chapter 1 Appendix

A   Figures

Pseudo Code, Figure A.1.

1. For draw $\widehat{\Pi}_{\mathtt{i}}$, compute both estimators $\widehat{\vartheta}_{\mathtt{i}}$ and $\widehat{\vartheta}_{\mathtt{i}}^*$ by plugging in to g^{-1}.

2. If both are outside of reals, the estimator is said to not exist for the given draw (2.05% of utilized sample).

3. If both estimators are in the reals, plot both in the Monte Carlo with equal weight 1 (44% of sample).

4. If one is in the reals and one is outside, plot only the one inside with weight of 2 (remainder of sample).

Pseudo Code, Figure A.2.

1. For draw $\widehat{\Pi}_{\mathtt{i}}$, compute both estimators $\widehat{\vartheta}_{\mathtt{i}}$ and $\widehat{\vartheta}_{\mathtt{i}}^*$ by plugging in to g^{-1}.

2. If both are outside of reals, the estimator is said to not exist for the given draw (2.05% of utilized sample).

3. If both estimators are in the reals

   (a) If one estimator yields a $\widehat{\tau}_{\mathtt{i}}$ which is closer to 2, plot only that estimator, with weight 2. If equidistant, plot both estimators, with weight 1 each.

Figure A.1: Distribution of An and Schorfheide model without means MLE at ϑ_0: T=250, N=10,000. Two-sided $(\alpha/2, 1-\alpha/2)$ confidence interval, $\alpha = 5\%$.
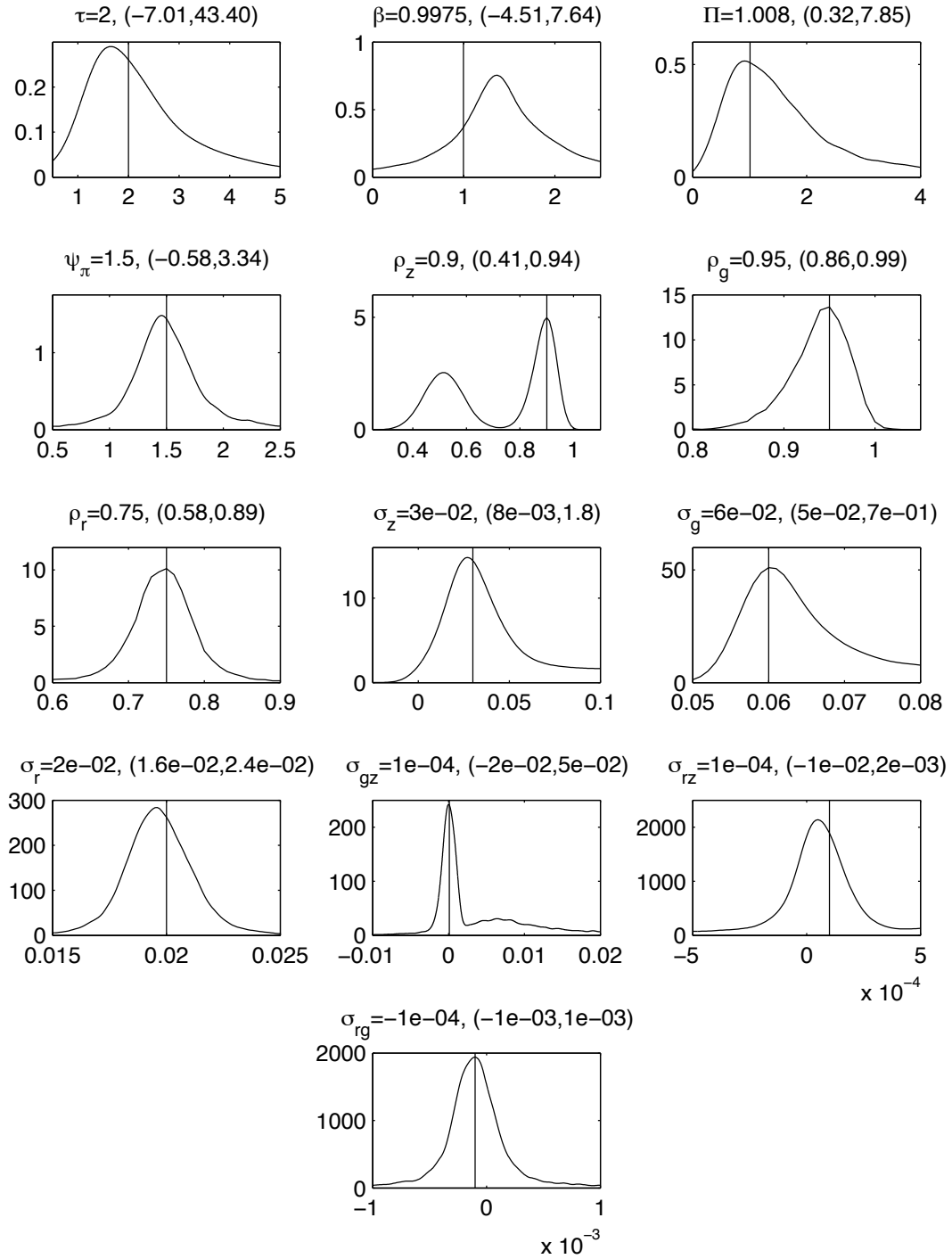
Figure A.2: Distribution of An and Schorfheide model without means MLE with $\tau \approx 2$ identification scheme ϑ_0: T=250, N=10,000. Two-sided $(\alpha/2, 1 - \alpha/2)$ confidence interval, $\alpha = 5\%$.
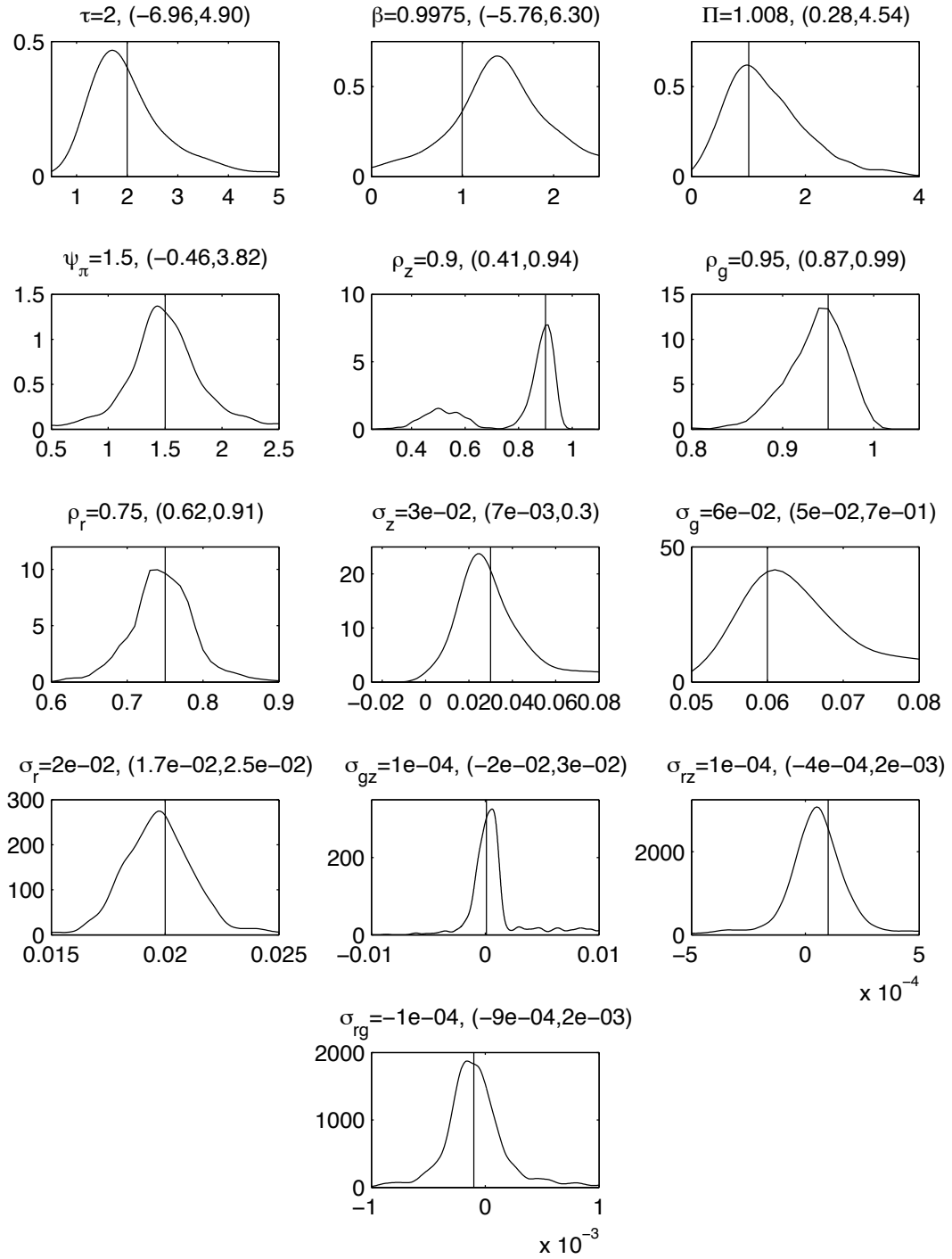
B   An and Schorfheide Model Appendix

The following small scale economy is similar to that presented in An and Schorfheide (2007). A distinction is that I will ultimately allow for statistical 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.

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.

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

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,

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

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

Representative Household. The representative household has real money bal-

ances, 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,

$$\beta E_t \frac{\lambda_{t+1}}{\lambda_t} = \frac{1}{\Pi_{t+1}} Q_{t+1|t} \tag{A.4}$$

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

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

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.

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

Partial Equilibrium Between Firms and Households. Firstly, plugging

Equation (A.7) into (A.3) gives

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

while combining Equations (A.4) and (A.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 (A.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

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

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

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

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:

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

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

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

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

Meanwhile, aggregate productivity follows

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

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

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

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

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 (A.1) does not matter, so that Φ_t may be written

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

Thus, using this definition of Φ_t and $F_t = \zeta_t Y_t$, the aggregate accounting equality Equation (A.15) may also be written

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

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 $\rho_{gz} = \sigma_{gz}/(\sigma_g\sigma_z)$, $\rho_{rz} = \sigma_{rz}/(\sigma_r\sigma_z)$, and $\rho_{rg} = \sigma_{rg}/(\sigma_r\sigma_g)$.

Nonlinear Equlibrium Equations and Solution. Ultimately, it will be useful to have the equilibrium equations expressed in log-levels of the endogenous variables for the pupose of solving the model in `Dynare`, but still ensuring direct compatibility of this solution with the log-linearized equilibrium equations elsewhere. Furthermore, many variables in the above equations can be eliminated by simple substitutions. First, simply rewriting Equation (A.8) gives

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

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 (A.13), Equation (A.9) may be written

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

Equation (A.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 (A.15) implies $\widehat{Y}_t^* = G_t\widehat{C}_t^*$. But also, when $\phi = 0$, Equation (A.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 (A.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^*:

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

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

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

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

$$-\ln Z_t + \rho_z \ln Z_{t-1} + \varepsilon_{zt} = 0 \tag{A.21}$$

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

In all, there are six key equilibrium equations (A.17) - (A.22) that completely characterize equilibrium for this model stated in terms of the six (detrended) variables of interest $\ln Z_t$, $\ln G_t$, $\ln R_t$, $\ln \widehat{Y}_t = \ln(Y_t/A_t)$, $\ln \Pi_t$, and $\ln \widehat{C}_t = \ln(C_t/A_t)$. The shocks are ε_{zt}, ε_{gt}, and ε_{rt}.

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

Table A.1: With-means An and Schorfheide model parameter and variable names.

		Structural Params (18)	Endogenous (6)		Innovations (3)	
1	τ	CRRA	z_t	Total factor prod.	ε_{zt}	To z_t
2	β	Discount factor	g_t	Gov spending	ε_{gt}	To g_t
3	ν	Inverse elas. of demnd	r_t	Nominal int rate	ε_{rt}	To r_t
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}	Covar of ε_{gt} and ε_{zt}				
17	σ_{rz}	Covar of ε_{rt} and ε_{zt}				
18	σ_{rg}	Covar of ε_{rt} and ε_{gt}				

C   An and Schorfheide Model With-Means

In the main text, it is assumed that the variables r_t, y_t, and π_t are logged deviations from the unconditional mean of each respective variable. However, the means of the data may also be useful for empirical analysis. Given the micro foundations of the model, it is natural to interpret the unconditional mean of each variable as the empirical analogue to the steady state. Specifically, under this assumption and using the notation introduced in Appendix B, and furthermore defining $Y_t = V_t - V(\theta)$, the VAR(1) in Equation (1.18) may be rewritten

$$
\underbrace{\ln \begin{bmatrix} R_t \\ \hat{Y}_t \\ \Pi_t \end{bmatrix}}_{V_t} - \underbrace{\ln \begin{bmatrix} R \\ \hat{Y} \\ \Pi \end{bmatrix}}_{V(\theta)} = \underbrace{\begin{bmatrix} \phi_{rr} & 0 & \phi_{r\pi} \\ \phi_{yr} & \rho_g & \phi_{y\pi} \\ \phi_{\pi r} & 0 & \phi_{\pi\pi} \end{bmatrix}}_{\Phi(\theta)} \left(\ln \begin{bmatrix} R_{t-1} \\ \hat{Y}_{t-1} \\ \Pi_{t-1} \end{bmatrix} - \ln \begin{bmatrix} R \\ \hat{Y} \\ \Pi \end{bmatrix} \right) + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \quad (A.23)
$$

where R_t is the with-mean nominal interest rate, $\hat{Y}_t$ is with-mean but detrended output, and Π_t is with-mean gross inflation. Meanwhile, in terms of the structural

Table A.2: With-means An and Schorfheide model candidate calibration θ_0 and parameter space Θ_0. ε=1e-6.

	Param	Lower	θ_0	Upper
1	τ	0.1	2	3.5
2	β	0.975	0.9975	1-ε
3	ν	ε	0.1	1
4	ϕ	50	53.68	60
5	γ	ε	0.5	$1-\varepsilon$
6	Π	$1+\varepsilon$	1.008	1.03
7	G	$1+\varepsilon$	1.18	1.25
8	ψ_π	-1	1.5	3
9	ψ_y	-1	0.125	1.25
10	ρ_z	ε	0.9	1-ε
11	ρ_g	ε	0.95	1-ε
12	ρ_r	ε	0.75	1-ε
13	σ_z	ε	3e-2	1
14	σ_g	ε	6e-2	1
15	σ_r	ε	2e-2	1
16	σ_{gz}	-1	1e-4	1
17	σ_{rz}	-1	1e-4	1
18	σ_{rg}	-1	-1e-4	1

parameters, $R = \gamma\Pi/\beta$ is the steady state of real interest rates, $\hat{Y} = G(1-\nu)^{1/\tau}$ is the steady state of detrended output, and Π, itself a structural parameter, is the steady state of inflation. Given these definitions, define a 3×1 vector $\Psi(\theta)$ by

$$
\underbrace{\begin{bmatrix} \psi_r \\ \psi_y \\ \psi_\pi \end{bmatrix}}_{\Psi(\theta)} = \underbrace{\begin{bmatrix} 1-\phi_{rr} & 0 & -\phi_{r\pi} \\ -\phi_{yr} & 1-\rho_g & -\phi_{y\pi} \\ -\phi_{\pi r} & 0 & 1-\phi_{\pi\pi} \end{bmatrix}}_{I_3-\Phi(\theta)} \ln \underbrace{\begin{bmatrix} \gamma\Pi/\beta \\ G(1-\nu)^{1/\tau} \\ \Pi \end{bmatrix}}_{V(\theta)}
$$

Then, given Equation (1.18), it is also the case that V_t – the vector of logged variables not separated from steady states – follows the rule of motion

$$
\underbrace{\ln \begin{bmatrix} R_t \\ \hat{Y}_t \\ \Pi_t \end{bmatrix}}_{V_t} = \underbrace{\begin{bmatrix} \psi_r \\ \psi_y \\ \psi_\pi \end{bmatrix}}_{\Psi(\theta)} + \underbrace{\begin{bmatrix} \phi_{rr} & 0 & \phi_{r\pi} \\ \phi_{yr} & \rho_g & \phi_{y\pi} \\ \phi_{\pi r} & 0 & \phi_{\pi\pi} \end{bmatrix}}_{\Phi(\theta)} \ln \begin{bmatrix} R_{t-1} \\ \hat{Y}_{t-1} \\ \Pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t} \tag{A.24}
$$

The unique reduced form parameters are collected in the 18×1 vector reduced form parameter

$$\underset{(18\times1)}{\Pi(\theta)} = (\psi_r, \psi_y, \psi_\pi, \phi_{rr}, \phi_{yr}, \phi_{\pi r}, \rho_g, \phi_{r\pi}, \phi_{y\pi}, \phi_{\pi\pi}, \omega_r, \omega_y, \omega_\pi, \omega_{yr}, \omega_{\pi r}, \omega_{\pi y})'$$

At θ_0, Equation (A.24) has the following realization; by definition, Φ and Ω are the same for the model posed with data separated from means, Equation (1.18).

$$\underbrace{\ln\begin{bmatrix} R_t \\ \hat{Y}_t \\ \Pi_t \end{bmatrix}}_{V_t} = \underbrace{\begin{bmatrix} -0.15 \\ 0.14 \\ 0.85 \end{bmatrix}}_{\Psi(\theta_0)} + \underbrace{\begin{bmatrix} 0.79 & 0 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0 & 0.62 \end{bmatrix}}_{\Phi(\theta_0)} \ln\begin{bmatrix} R_{t-1} \\ \hat{Y}_{t-1} \\ \Pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{rt} \\ u_{yt} \\ u_{\pi t} \end{bmatrix}}_{U_t}$$

$$\Omega(\theta_0) = (1e-4) \times \begin{bmatrix} 6 & \cdot & \cdot \\ 7 & 58 & \cdot \\ 7 & 21 & 20 \end{bmatrix}$$

Since Ψ may be estimated, there are in this case $n_\Pi = 16$ reduced form parameters. So, now potentially 16 of the $n_\theta = 18$ structural parameters are identifiable, meaning, up to 3 of the 5 parameters that were previously set can now be estimated. A figure depicting the functional dependence of each element of Π on θ is given in Figure A.3, which I now use to choose which two structural parameters must be set to constants for their complement to be conditionally identified. Firstly, consider both γ and G. Since these are not even included in the parameters Φ and Ω, they were not previously identifiable. However, Ψ is a function of both, so consider letting both be free. Second, the two parameters ν and ϕ were previously set due to their linear dependence in κ. However, Ψ is a function of both, so consider letting ν be estimated. This results in the conditional identification scheme $\overline{\alpha} = (\overline{\phi}, \overline{\psi}_\pi)$ so that

$$\underset{(16\times1)}{\vartheta} = (\tau, \beta, \nu, \gamma, \Pi, G, \psi_\pi, \rho_z, \rho_g, \rho_r, \sigma_z, \sigma_g, \sigma_r, \sigma_{gz}, \sigma_{rz}, \sigma_{rg})'$$

Since $n_\vartheta = 16 = n_\Pi$, if ϑ is identified at a point, it is exactly identified. Indeed, the

Π / θ	1 τ	2 β	3 ν	4 ϕ	5 γ	6 Π	7 G	8 ψ_π	9 ψ_y	10 ρ_z	11 ρ_g	12 ρ_r	13 σ_z	14 σ_g	15 σ_r	16 σ_{gz}	17 σ_{rz}	18 σ_{rg}
1 ψ_r	×	×	×	×	×	×		×	×	×		×						
2 ψ_y	×	×	×	×	×	×	×	×	×	×	×	×						
3 ψ_π	×	×	×	×	×	×		×	×	×		×						
4 ϕ_{rr}	×	×	×	×		×		×	×	×		×						
5 ϕ_{yr}	×	×	×	×		×		×	×	×	×	×						
6 $\phi_{\pi r}$	×	×	×	×		×		×	×	×		×						
7 ρ_g											×							
8 $\phi_{r\pi}$	×	×	×	×		×		×	×	×		×						
9 $\phi_{y\pi}$	×	×	×	×		×		×	×	×	×	×						
10 $\phi_{\pi\pi}$	×	×	×	×		×		×	×	×		×						
11 ω_r	×	×	×	×		×		×	×	×		×	×		×		×	
12 ω_y	×	×	×	×		×		×	×	×		×	×	×	×	×	×	×
13 ω_π	×	×	×	×		×		×	×	×		×	×		×		×	
14 ω_{yr}	×	×	×	×		×		×	×	×		×	×		×	×	×	×
15 $\omega_{\pi r}$	×	×	×	×		×		×	×	×		×	×		×		×	
16 $\omega_{\pi y}$	×	×	×	×		×		×	×	×		×	×		×	×	×	×

Figure A.3: An and Schorfheide model with–means functional mapping $g : \theta \to \Pi$.

Table A.3: Identification of ϑ in with-means model: 100,000 uniformly chosen pts from Θ_0.

	Yes	No
Locally Identified	99.98 %	0.02 %
Globally Identified in Θ_0	92.6%	7.4 %
Globally Identified in All $\Theta \subset \mathbb{R}^{16}$	30%	70%

16×16 Jacobian $J(\vartheta_0; \overline{\alpha}) = \partial\Pi(\vartheta; \overline{\alpha})/\partial\vartheta'|_{\vartheta=\vartheta_0}$, is full column rank, thus satisfying a necessary and sufficient condition for ϑ to be locally identified at ϑ_0. Thus, it seems that adding the means of the data into estimation has allowed the estimation of three more structural parameters. Such claims that adding more aspects of the data results in the identification of more structural parameters seem intuitive, and are frequently made. Yet, global identification has not yet been verified.

As in the previous two examples, the inverse mapping g^{-1} is calculated analytically and there are exactly two solutions. The values corresponding to ϑ_0 are given in Figure A.4. As in the model not using means, ϑ_0 is globally identified in the reals, $\mathbb{R}^{16}$. However, again there are points where the bounds of Θ_0 may not be used to distinguish between two points. One example is ϑ_1 and ϑ_1^*, listed in Figure A.5. Once again, for both ϑ_0 and ϑ_1 it is demonstrated that the impulse responses, besides those corresponding to ε_{gt}, differ for observationally equivalent points, although the magnitude of the difference between ϑ_1 and ϑ_1^* is again small. Thus, identifying restrictions are not always normalizations with respect to the impulse responses.

Now consider the identification of ϑ at other realizations in Θ. Both local and global identification are summarized in Table A.3. As before, nearly all points are locally identified. However, there is now a relatively larger portion of points, 7.4%, which are not globally identified in Θ_0. These require creative identifying restrictions.

Finally, bootstrapped confidence intervals are computed in Figures A.6 and A.7 using no identifying restriction, and the better τ restriction, respectively. Once again, the most obvious distinction is in the size of the confidence interval for $\widehat{\tau}$. A long right-tail under no identifying restrion is much smaller when the restriction is accounted for. Pseudo-code resembles that for the model without-means.

Figure A.4: An and Schorfheide with means observational equivalence in $\mathbb{C}^{16}$ but not Θ_0, and impulse-responses: ϑ_0 ($\circ$) vs ϑ_0^* ($+$). $\Pi = g(x; \overline{\alpha})$; $x = \vartheta_0$ or ϑ_0^*.

		$g(x, \overline{\alpha})$		ϑ_0	ϑ_0^*
1	ψ_r	-0.14	$\boldsymbol{\tau}$	2	**-45.4**
2	ψ_y	0.14	β	0.9975	0.9975
3	ψ_π	0.08	$\boldsymbol{\nu}$	0.1	**1.66**
4	ϕ_{rr}	0.79	γ	0.5	0.5
5	ϕ_{yr}	0.19	Π	1.008	1.008
6	$\phi_{\pi r}$	0.12	$\boldsymbol{G}$	1.18	**1.11+0.08i**
7	ρ_g	0.95	ψ_π	1.5	1.5
8	$\phi_{r\pi}$	0.25	$\boldsymbol{\rho_z}$	0.9	**0.51**
9	$\phi_{y\pi}$	-0.46	ρ_g	0.95	0.95
10	$\phi_{\pi\pi}$	0.62	ρ_r	0.75	0.75
11	ω_r	2e-2	$\boldsymbol{\sigma_z}$	3e-2	**1.85**
12	ω_y	7e-2	σ_g	6e-2	6e-2
13	ω_π	4e-2	σ_r	2e-2	2e-2
14	ω_{yr}	5e-4	$\boldsymbol{\sigma_{gz}}$	1e-4	**-1e2**
15	$\omega_{\pi r}$	7e-4	$\boldsymbol{\sigma_{rz}}$	1e-4	**1e2**
16	$\omega_{\pi y}$	23e-4	σ_{rg}	-1e-4	-1e-4
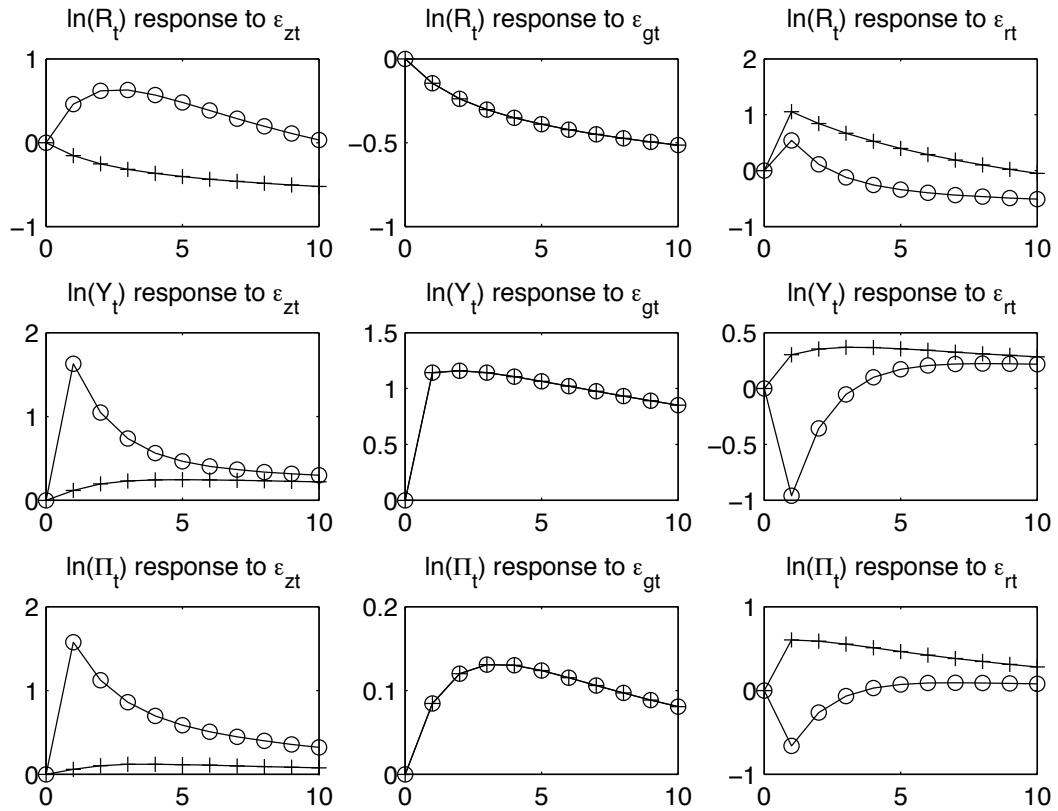
Figure A.5: An and Schorfheide with means observational equivalence in Θ_0, and impulse-responses: ϑ_0 ($\circ$) vs ϑ_0^* (+). $\Pi = g(x; \overline{\alpha})$; $x = \vartheta_1$ or ϑ_1^*.

		$g(x, \overline{\alpha})$		ϑ_1	ϑ_1^*
1	ψ_r	-0.09	$\boldsymbol{\tau}$	1.59	**2.06**
2	ψ_y	-0.006	β	0.9815	0.9815
3	ψ_π	0.001	$\boldsymbol{\nu}$	0.23	**0.28**
4	ϕ_{rr}	0.31	γ	0.86	0.86
5	ϕ_{yr}	-0.06	Π	1.001	1.001
6	$\phi_{\pi r}$	-0.004	$\boldsymbol{G}$	1.11	**1.10**
7	ρ_g	0.91	ψ_π	1.72	1.72
8	$\phi_{r\pi}$	0.42	$\boldsymbol{\rho_z}$	0.27	**0.26**
9	$\phi_{y\pi}$	-5.47	ρ_g	0.91	0.91
10	$\phi_{\pi\pi}$	0.23	ρ_r	0.32	0.32
11	ω_r	0.67	$\boldsymbol{\sigma_z}$	0.73	**0.37**
12	ω_y	0.99	σ_g	0.69	0.69
13	ω_π	0.06	σ_r	0.78	0.78
14	ω_{yr}	-0.59	$\boldsymbol{\sigma_{gz}}$	-0.42	**-0.23**
15	$\omega_{\pi r}$	-0.04	$\boldsymbol{\sigma_{rz}}$	0.53	**0.11**
16	$\omega_{\pi y}$	0.05	σ_{rg}	-0.34	-0.34
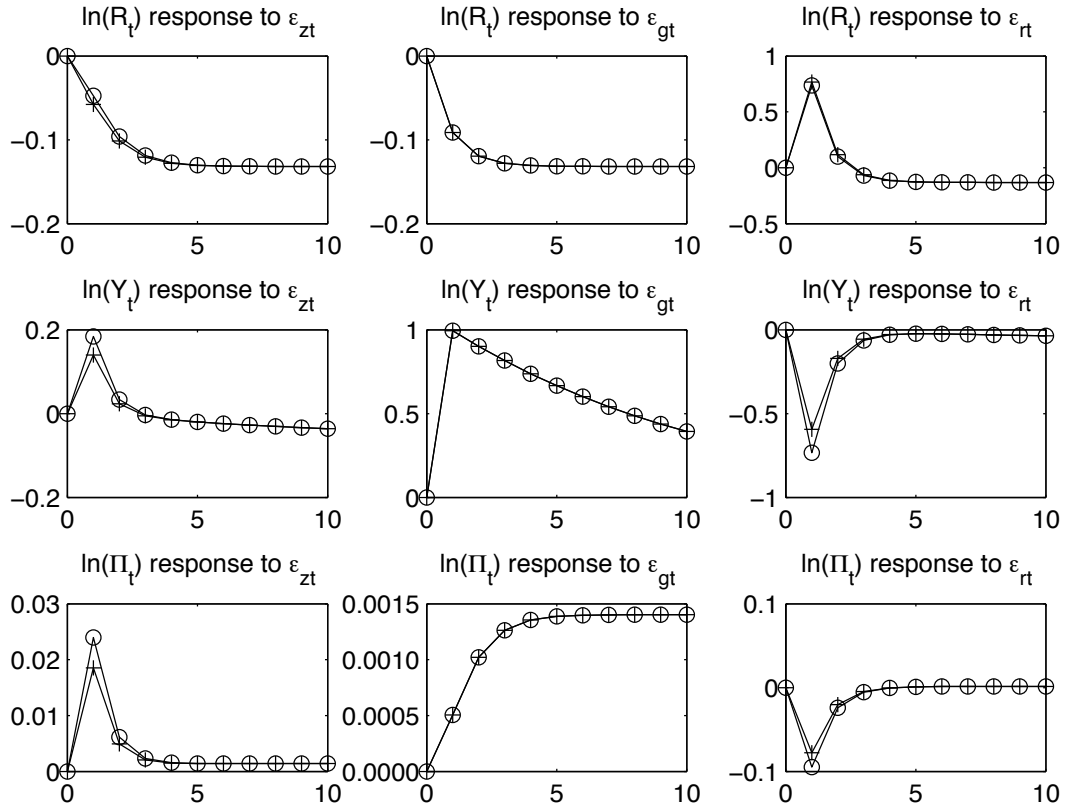
Figure A.6: Distribution of An and Schorfheide with means MLE at ϑ_0: T=250, N=10,000.
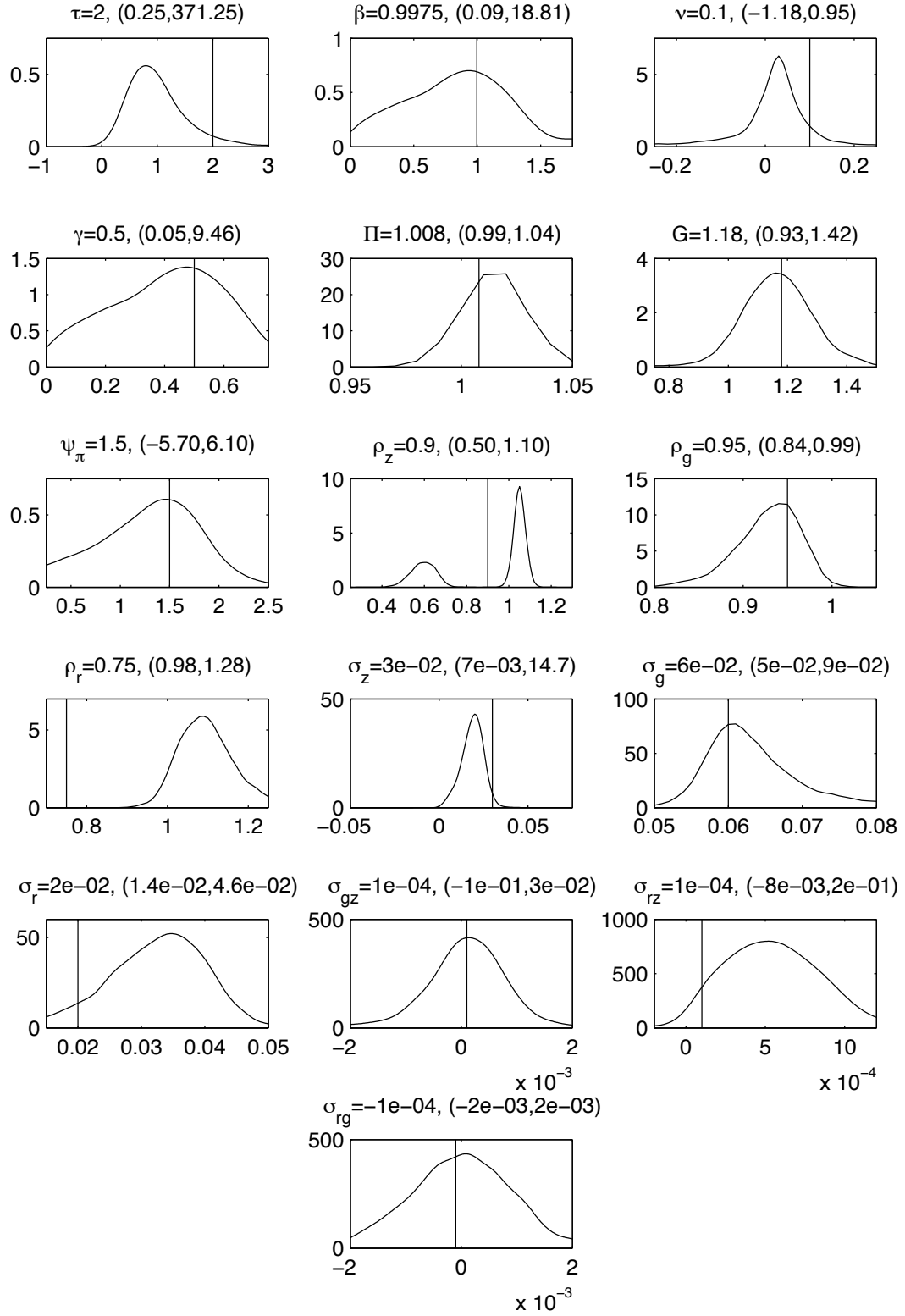
Figure A.7: Distribution of An and Schorfheide with means MLE with $\tau \approx 2$ identification.
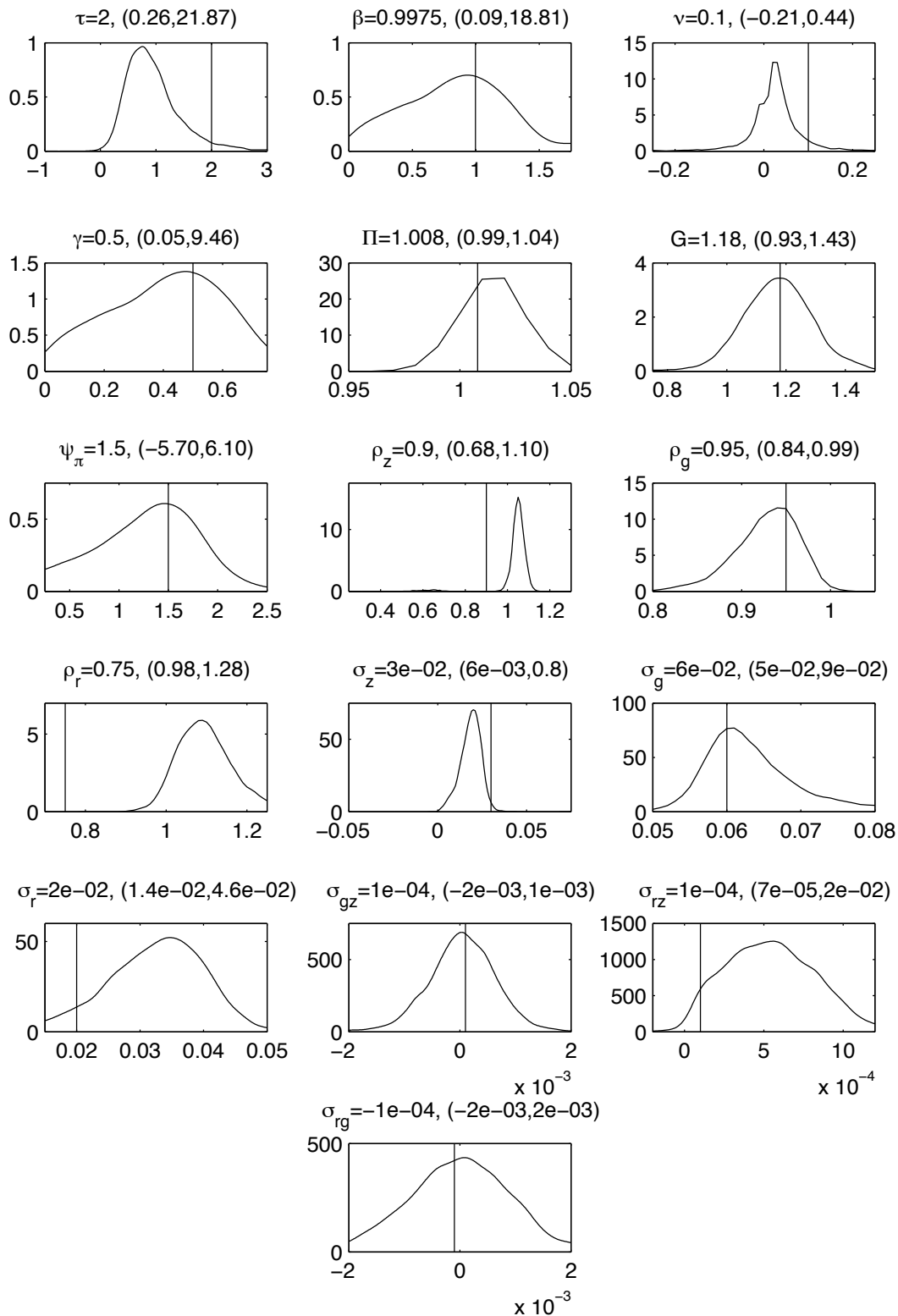
D Smets and Wouters (2007) Appendix

This section presents the linearized equilibrium conditions of Smets and Wouters (2007). References to where these equations may be verified include the Appendix and code for the paper, appearing on the *American Economic Review* website.

Any apparent deviations between the expressions used here and Smets and Wouters's original paper are only the result of benign simplifications and notational differences. In terms of notation, the conventions used in the derivation of the An and Schorfheide (2007) model above – capital letters for variables, capitals with no subscript t for steady states, hats for detrended variables, and lower case for log deviation from steady states – are preserved here for ease of comparison between the two models, and context for how this much more elaborate model is derived. Specific names of variables and parameters are given in Table A.4. Candidate parameter values θ_0 and parameter space Θ are given in Table A.5.

Equilibrium With Real Rigidities. The Smets and Wouters model is composed of two separate equilibria. The first, corresponding to the theoretical setting in which prices and wages are sticky, is represented by 14 unique equations paired with 14 endogenous variables.

The first difference separating the Smets and Wouters from An and Schorfheide model is the presence of a rule of motion for capital. Nominal, detrended capital – also known as "installed" capital – follows the rule of motion

$$k_t = \frac{1-\delta}{\gamma}k_{t-1} + \frac{1}{\gamma}\frac{\widehat{I}}{\widehat{K}}i_t + s\gamma(1+\beta\gamma^{1-\tau})\frac{\widehat{I}}{\widehat{K}}e_t \tag{A.25}$$

where $\widehat{I}/\widehat{K} = \gamma - (1-\delta)$ is the ratio of steady state investment to capital. However, there are similarities between Smets and Wouters from An and Schorfheide. The Euler equation arising from household optimization may be written as follows. For intuition for how it is derived, note the similarity between this expression and the aggregate demand equality from the AS model. Here, c_t is the log deviation of

Table A.4: Elements of Smets and Wouters (2007). Normalized notation with An and Schorfheide model.

		Structural Params (41)		**Endogenous** (33)		**Errors** (7)
1	τ	CRRA	z_t	Total factor prod.	ε_{zt}	To z_t
2	ℓ	Disutility of labor	g_t	Gov. spending	ε_{gt}	To g_t
3	β	Discount factor	e_t	Price of inv. v. cons.	ε_{et}	To e_t
4	h	Habit formation in cons.	b_t	Bond premium	ε_{bt}	To b_t
5	s	Investment adj. cost	f_t	Fed shock int. rate	ε_{ft}	To f_t
6	α	Capital inten. in prod.	m_t^p	Shock price markup	ε_{pt}	To m_t^p
7	γ	Avg. gr. rate of prod.	m_t^w	Shock wage markup	ε_{wt}	To m_t^w
8	Φ	Fixed cost of prod.	λ_t^p	Func. of m_t^p and ε_t^p		
9	δ	Depreciation of capital	λ_t^w	Func. of m_t^w and ε_t^w		
10	u	Capital util. inten.	k_t	Installed Capital		
11	ι_p	Price indexation	c_t	Real consumption		
12	ξ_p	Calvo price prob.	i_t	Real investment		
13	ζ_p	Kimball price agg. curv.	π_t	Inflation		
14	ι_w	Wage indexation	w_t	Real wage		
15	ξ_w	Calvo wage prob.	μ_t^p	Price markup		
16	ζ_w	Kimball wage agg. curv.	μ_t^w	Wage markup		
17	λ_w	Wage markup	q_t	Tobin's Q		
18	$\widehat{Y}$	St. state detr. output	r_t	Nominal int. rate		
19	Π	St. state inflation	r_t^k	Real rent. rate on k_t		
20	G	St. state gov. spending	s_t	Utilized captial		
21	ψ_π	Taylor rule inflation	u_t	Capacity utilization		
22	ψ_y	Taylor rule output	l_t	Labor hours		
23	ψ_Δ	Tay. rule output chg.	y_t	Real output		
24	ρ_r	r_t persistence	k_t^*	Nat. level inst. cap.		
25	ρ_b	b_t persistence	c_t^*	Nat. real consump.		
26	ρ_e	e_t persistence	i_t^*	Nat. real invest.		
27	ρ_f	f_t persistence	w_t^*	Nat. real wage		
28	ρ_z	z_t persistence	q_t^*	Nat. Tobin's Q		
29	ρ_g	g_t persistence	r_t^*	Nat. nom. int. rate		
30	ρ_p	m_t^p persistence	r_t^{k*}	Nat. real r.r. on k_t^*		
31	ρ_w	m_t^w persistence	s_t^*	Nat. utilized cap.		
32	ϑ_{gz}	ε_{zt} coeff for g_t AR(1)	l_t^*	Nat. labor hrs.		
33	ϑ_p	MA(1) coeff for m_t^p	y_t^*	Nat. real output		
34	ϑ_w	MA(1) coeff for m_t^w				
35	σ_b	ε_{bt} std error				
36	σ_e	ε_{et} std error				
37	σ_f	ε_{ft} std error				
38	σ_z	ε_{zt} std error				
39	σ_g	ε_{gt} std error				
40	σ_p	ε_{pt} std error				
41	σ_w	ε_{wt} std error				

Table A.5: Smets and Wouters calibration θ_0 and parameter space Θ. ε=1e-6.

	Param	**Lower**	$\boldsymbol{\theta_0}$	**Upper**
1	τ	0.1	0.5	3.5
2	ℓ	0.1	2	3.5
3	β	0.975	0.9975	1-ε
4	h	0	0.7	1-ε
5	s	0	4	10
6	α	ε	0.3	$1-\varepsilon$
7	γ	ε	0.95	$1-\varepsilon$
8	Φ	0	1.25	5
9	δ	ε	0.1	$1-\varepsilon$
10	u	ε	0.5	$1-\varepsilon$
11	ι_p	ε	0.5	$1-\varepsilon$
12	ξ_p	ε	0.5	$1-\varepsilon$
13	ζ_p	5	10	15
14	ι_w	ε	0.5	$1-\varepsilon$
15	ξ_w	ε	0.5	$1-\varepsilon$
16	ζ_w	5	10	15
17	λ_w	ε	1.5	5
18	$\widehat{Y}$	$1+\varepsilon$	1.065	1.2
19	Π	$1+\varepsilon$	1.008	1.2
20	G	$1+\varepsilon$	1.18	1.2
21	ψ_π	-1	1.5	3
22	ψ_y	-1	0.125	1.25
23	ψ_Δ	-1	0.125	1.25
24	ρ_r	ε	0.75	$1-\varepsilon$
25	ρ_z	ε	0.9	$1-\varepsilon$
26	ρ_g	ε	0.95	$1-\varepsilon$
27	ρ_e	ε	0.5	$1-\varepsilon$
28	ρ_b	ε	0.5	$1-\varepsilon$
29	ρ_f	ε	0.5	$1-\varepsilon$
30	ρ_p	ε	0.5	$1-\varepsilon$
31	ρ_w	ε	0.5	$1-\varepsilon$
32	ϑ_{gz}	ε	0.5	$1-\varepsilon$
33	ϑ_p	ε	0.5	$1-\varepsilon$
34	ϑ_w	ε	0.5	$1-\varepsilon$
35	σ_b	ε	1e-2	$1-\varepsilon$
36	σ_e	ε	1e-2	$1-\varepsilon$
37	σ_f	ε	1e-2	$1-\varepsilon$
38	σ_z	ε	3e-2	$1-\varepsilon$
39	σ_g	ε	6e-2	$1-\varepsilon$
40	σ_p	ε	1e-2	$1-\varepsilon$
41	σ_w	ε	1e-2	$1-\varepsilon$

real, detrended consumption from its natural rate.

$$c_t = \frac{\gamma}{\gamma + h} E_t c_{t+1} + \frac{h}{\gamma + h} c_{t-1} - \frac{1-\tau}{\tau} \frac{\gamma}{\gamma + h} \kappa_e \left(l_t - E_t l_{t+1}\right)$$

$$- \frac{1}{\tau} \frac{\gamma - h}{\gamma + h} \left(r_t - E_t \pi_{t+1} + b_t\right) \quad \text{(A.26)}$$

where

$$\kappa_e = \frac{1}{1 + \lambda_w} \frac{1-\alpha}{\alpha} R^K \frac{\widehat{K}}{\widehat{Y}} \frac{\widehat{Y}}{\widehat{C}}$$

$$\frac{\widehat{Y}}{\widehat{C}} = \frac{1}{1 - \frac{G}{\widehat{Y}} - \frac{\widehat{I}}{\widehat{K}} \frac{\widehat{K}}{\widehat{Y}}} \qquad \frac{\widehat{I}}{\widehat{K}} = \gamma - (1-\delta) \qquad \frac{\widehat{K}}{\widehat{Y}} = \frac{\widehat{Y} + \Phi}{\widehat{Y}} \left(\frac{L}{\widehat{K}}\right)^{\alpha-1} \qquad \frac{L}{\widehat{K}} = \frac{1-\alpha}{\alpha} \frac{R^K}{\widehat{W}}$$

$$\widehat{W} = \left(\frac{\widehat{Y}}{\widehat{Y} + \Phi} \frac{\alpha^\alpha (1-\alpha)^{1-\alpha}}{(R^K)^\alpha}\right)^{1/(1-\alpha)} \qquad R^K = \frac{\gamma^\tau}{\beta} - (1-\delta)$$

W is the steady state of detrended real wages and R^K is the steady state of the rental rate on depreciable capital. The parameters $(\widehat{Y}, G)$ are steady state values that appear in the structural parameter θ. $(\widehat{K}, \widehat{C}, \widehat{I}, L)$ are steady state parameters which are functions of the structural parameters, as described below. As in the derivation of the AS model previously, hats indicate which steady state parameters correspond to detrended variables. To reiterate, the definitions of all structural parameters and variables are given in Table A.4.

Next, we have a relationship between investment and Tobin's Q.

$$i_t = \frac{\beta \gamma^{1-\tau}}{1 + \beta \gamma^{1-\tau}} E_t i_{t+1} + \frac{1}{1 + \beta \gamma^{1-\tau}} i_{t-1} + \frac{1}{s\gamma^2} \frac{1}{1 + \beta \gamma^{1-\tau}} q_t + e_t \quad \text{(A.27)}$$

Investment is real and detrended. Meanwhile, the Phillips curve is

$$\pi_t = \frac{\beta \gamma^{1-\tau}}{1 + \iota_p \beta \gamma^{1-\tau}} E_t \pi_{t+1} + \frac{\iota_p}{1 + \iota_p \beta \gamma^{1-\tau}} \pi_{t-1} - \kappa_p \mu_t^p + m_t^p \quad \text{(A.28)}$$

where

$$\kappa_p = \frac{(1 - \xi_p \beta \gamma^{1-\tau})(1 - \xi_p)}{\xi_p (1 + \zeta_p(\Phi/\widehat{Y}))(1 + \iota_p \beta \gamma^{1-\tau})}$$

and the wage relation is

$$w_t = \frac{\beta\gamma^{1-\tau}}{1+\beta\gamma^{1-\tau}}\left(E_t w_{t+1} + E_t \pi_{t+1}\right) + \frac{1}{\beta\gamma^{1-\tau}}(w_{t-1} + \iota_w \pi_{t-1})$$
$$- \frac{1+\iota_w\beta\gamma^{1-\tau}}{1+\beta\gamma^{1-\tau}}\pi_t - \kappa_w\mu_t^w + m_t^w \quad \text{(A.29)}$$

where

$$\kappa_w = \frac{(1-\xi_w\beta\gamma^{1-\tau})(1-\xi_w)}{\xi_w(1+\zeta_w\lambda_w)(1+\iota_w\beta\gamma^{1-\tau})}$$

Wages are real and detrended. μ_t^p and μ_t^w are the markups to prices and wages, respectively, defined by

$$\mu_t^p = \alpha(s_t - l_t) - w_t + z_t \quad \text{(A.30)}$$

$$\mu_t^w = w_t - \ell l_t - \frac{\gamma}{\gamma-h}c_t + \frac{h}{\gamma-h}c_{t-1} \quad \text{(A.31)}$$

where markups are nominal and not detrended. The same is true for Tobin's Q, which follows the rule of motion

$$q_t = \beta(1-\delta)\gamma^{-\tau}E_t q_{t+1} + E_t\pi_{t+1} + (1 - \beta(1-\delta)\gamma^{-\tau})E_t r_{t+1}^k - r_t - b_t \quad \text{(A.32)}$$

Taylor's rule for the nominal interest rate is written as follows. Starred variables are those arising from the equilibrium derived without real rigidities, to be expressed in the following subsection; in particular, y_t^* will be known as the "natural" rate of output. f_t is the idiosyncratic component of Fed policy not captured by the Taylor rule.

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_\pi\pi_t + ((1-\rho_r)\psi_y + \psi_\Delta)(y_t - y_t^*)$$
$$- \psi_\Delta\left(y_{t-1} - y_{t-1}^*\right) + f_t \quad \text{(A.33)}$$

while the real but not detrended rental rate for installed capital k_t is given by

$$r_t^k = l_t + w_t - s_t \quad \text{(A.34)}$$

where utilized capital s_t is related to installed capital (both nominal, detrended)

by

$$s_t = k_{t-1} + u_t \tag{A.35}$$

where u_t is nominal and not detrended capacity utilization, simply defined by

$$u_t = \frac{1-u}{u} r_t^k \tag{A.36}$$

Nominal and not detrended labor hours are given by

$$l_t = \frac{1}{1-\alpha} \frac{\widehat{Y}}{\widehat{Y} + \Phi} y_t - \frac{\alpha}{1-\alpha} s_t - \frac{1}{1-\alpha} z_t \tag{A.37}$$

and finally, real and detrended output is defined by the aggregate accounting equality.

$$y_t = \frac{\widehat{C}}{\widehat{Y}} c_t + \frac{\widehat{I}}{\widehat{K}} \frac{\widehat{K}}{\widehat{Y}} i_t + R^K \frac{\widehat{K}}{\widehat{Y}} \frac{1-u}{u} r_t^k + g_t \tag{A.38}$$

The 14 equations Equations (A.26)-(A.38) define rules of motion for the 14 endogenous variables k_t, c_t, i_t, π_t, w_t, μ_t^p, μ_t^w, q_t, r_t, r_t^k, s_t, u_t, l_t, and y_t, in that order. The next section defines a rule of motion for the natural rate of output, y_t^*, which appeared in the Taylor rule. The remaining 7 variables used in the above equations, z_t, g_t, e_t, b_t, f_t, m_t^p, and m_t^w, will be given reduced form rules of motion in the section after next.

Equilibrium Without Real Rigidities. The second set of equilibrium conditions defining the SW model correspond to the theoretical setting in which prices and wages are flexible. This set of equilibrium conditions is reduced to 10 unique equations paired with 10 endogenous variables. Keeping with the setting of the SW model, I shall refer to all of these variables as *natural* rates and levels. The natural level of installed capital is defined by

$$k_t^* = \frac{1-\delta}{\gamma} k_{t-1}^* + \frac{1}{\gamma} \frac{\widehat{I}}{\widehat{K}} i_t^* + s\gamma(1 + \beta\gamma^{1-\tau}) \frac{\widehat{I}}{\widehat{K}} e_t \tag{A.39}$$

while the natural rate of consumption c_t^* follows

$$c_t^* = -\frac{\gamma - h}{\gamma}w_t^* - \frac{\gamma - h}{\gamma}\ell l_t^* - \frac{h}{\gamma}c_{t-1}^* \tag{A.40}$$

Notice, this expression for c_t^* is much different than the consumption Euler equation for c_t in Equation (A.26). This is because the flexible price/wage analogue of Equation (A.26) will be used to define a rule of motion for r_t^* in the absence of a flexible price/wage Taylor rule. The natural rate of investment follows

$$i_t^* = \frac{\beta\gamma^{1-\tau}}{1 + \beta\gamma^{1-\tau}}E_t i_{t+1}^* + \frac{1}{1 + \beta\gamma^{1-\tau}}i_{t-1}^* + \frac{1}{s\gamma^2}\frac{1}{1 + \beta\gamma^{1-\tau}}q_t^* + e_t \tag{A.41}$$

Given that price markup is zero in this case,

$$w_t^* = \alpha(s_t^* - l_t^*) + z_t \tag{A.42}$$

while the natural Tobin's Q is

$$q_t^* = \beta(1-\delta)\gamma^{-\tau}E_t q_{t+1}^* + (1 - \beta(1-\delta)\gamma^{-\tau})E_t r_{t+1}^{k*} - r_t^* - b_t \tag{A.43}$$

and as explained, in place of a Taylor rule, r_t^* is defined by a rearranged consumption Euler equation as

$$r_t^* = -\frac{\tau}{\gamma - h}\left(c_t^* - \gamma E_t c_{t+1}^* - hc_{t-1}^*\right) - (1 - \tau)\frac{\gamma}{\gamma - h}\kappa_e(l_t^* - E_t l_{t+1}^*) - b_t \tag{A.44}$$

where κ_e is defined following Equation (A.26), and the natural rental rate for installed capital k_t^* is

$$r_t^{k*} = l_t^* + w_t^* - s_t^* \tag{A.45}$$

and natural utilized capital s_t^* is related to the natural level of installed capital by

$$s_t^* = k_{t-1}^* + \frac{1-u}{u}r_t^{k*} \tag{A.46}$$

Finally, natural labor hours and the natural output are, respectively,

$$l_t^* = \frac{1}{1-\alpha} \frac{\widehat{Y}}{\widehat{Y} + \Phi} y_t^* - \frac{\alpha}{1-\alpha} s_t^* - \frac{1}{1-\alpha} z_t \tag{A.47}$$

and

$$y_t^* = \frac{\widehat{C}}{\widehat{Y}} c_t^* + \frac{\widehat{I}}{\widehat{K}} \frac{\widehat{K}}{\widehat{Y}} i_t^* + R^K \frac{\widehat{K}}{\widehat{Y}} \frac{1-u}{u} r_t^{k*} + g_t \tag{A.48}$$

To conclude, the 10 numbered equations (A.40)-(A.48) define rules of motion for the 10 natural rates and levels k_t^*, c_t^*, i_t^*, w_t^*, q_t^*, r_t^*, r_t^{k*}, s_t^*, l_t^*, and y_t^*. Now we move on to the remaining 7 variables z_t, g_t, e_t, b_t, f_t, m_t^p, and m_t^w.

Reduced Form Processes. Seven of the variables used above are not defined by equilibrium conditions. Instead, five of these are AR(1) and the last two ARMA(1,1).

$$z_t = \rho_z z_{t-1} + \varepsilon_{zt} \tag{A.49}$$

$$g_t = \rho_g g_{t-1} + \varepsilon_{gt} + \vartheta_{gz}\varepsilon_{zt} \tag{A.50}$$

$$e_t = \rho_e e_{t-1} + \varepsilon_{et} \tag{A.51}$$

$$b_t = \rho_b b_{t-1} + \varepsilon_{bt} \tag{A.52}$$

$$f_t = \rho_f f_{t-1} + \varepsilon_{ft} \tag{A.53}$$

$$m_t^p = \rho_p m_{t-1}^p + \varepsilon_{pt} - \vartheta_p \varepsilon_{pt-1} \tag{A.54}$$

$$m_t^w = \rho_w m_{t-1}^w + \varepsilon_{wt} - \vartheta_w \varepsilon_{wt-1} \tag{A.55}$$

All innovations are iid. For the purposes of this paper, it will be convenient to write each of the ARMA(1,1) processes Equations (A.54) and (A.55) and 2-dimensional VAR(1)'s. Specifically, defining

$$\lambda_t^p = \rho_p m_t^p - \vartheta_p \varepsilon_{pt}$$

$$\lambda_t^w = \rho_w m_t^w - \vartheta_w \varepsilon_{wt}$$

Then it is easy to verify that Equations (A.54) and (A.55) may equivalently be represented by the set of four equations

$$m_t^p = \lambda_t^p + \varepsilon_{pt} \tag{A.56}$$

$$m_t^w = \lambda_t^w + \varepsilon_{wt} \tag{A.57}$$

$$\lambda_t^p = \rho_p \lambda_{t-1}^p + (\rho_p - \vartheta_p)\varepsilon_{pt} \tag{A.58}$$

$$\lambda_t^w = \rho_w \lambda_{t-1}^w + (\rho_w - \vartheta_w)\varepsilon_{wt} \tag{A.59}$$

Steady State Conditions. The steady states of output, inflation, and government spending, $\widehat{Y}$, Π, and G, are included in the structural parameters, and $R = (\Pi\gamma^\tau)/\beta$. Using the definitions following Equation (A.26), we already have explicit functions for the steady states of the rental rate R^K and wage W. From those definitions we can also say that

$$\widehat{K} = \frac{\widehat{I}}{\gamma - (1-\delta)} \qquad \widehat{C} = \left(1 - \frac{G}{\widehat{Y}} - \frac{\widehat{I}}{\widehat{K}}\frac{\widehat{K}}{\widehat{Y}}\right)\widehat{Y}$$

$$\widehat{I} = (\gamma - (1-\delta))\frac{\widehat{K}}{\widehat{Y}}\widehat{Y} \qquad L = \frac{1-\alpha}{\alpha}\frac{R^K}{\widehat{W}}\widehat{K}$$

VARMA(3,2) Representation. The ABCD representation of this model is given in Figure A.8. The 7 states having exogenously defined rules of motion are

$$X_{1t} = [z_t, g_t, e_t, b_t, f_t, \lambda_{pt}, \lambda_{wt}]'$$

The 4 states corresponding to the equilibrium derived without real rigidities are

$$X_{2t} = [k_t^*, c_t^*, i_t^*, y_t^*]'$$

Te 7 states corresponding to the equilibrium derived with real rigidities are

$$X_{3t} = [k_t, c_t, i_t, \pi_t, w_t, r_t, y_t]'$$

$$
\underbrace{\begin{bmatrix} X_{1t} \\ {\scriptstyle 7\times1} \\ X_{2t} \\ {\scriptstyle 4\times1} \\ X_{3t} \\ {\scriptstyle 7\times1} \end{bmatrix}}_{\substack{x_t \\ 18\times1}}
=
\underbrace{\begin{bmatrix}
\underbrace{\begin{bmatrix} \rho_z & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \rho_w \end{bmatrix}}_{7\times7} & \underbrace{0}_{7\times4} & \underbrace{0}_{7\times7} \\
\begin{bmatrix} A^1_{21} \\ {\scriptstyle 4\times4} \end{bmatrix} \begin{bmatrix} 0 \\ {\scriptstyle 4\times3} \end{bmatrix} & \begin{bmatrix} A^1_{22} \\ {\scriptstyle 4\times3} \end{bmatrix} \begin{bmatrix} 0 \\ {\scriptstyle 4\times1} \end{bmatrix} & \begin{bmatrix} 0 \\ {\scriptstyle 4\times7} \end{bmatrix} \\
\begin{bmatrix} A^1_{31} \\ {\scriptstyle 1\times7} \\ C_{21} \end{bmatrix} & \begin{bmatrix} A^1_{32} \\ {\scriptstyle 1\times4} \\ C_{21} \end{bmatrix} & \begin{bmatrix} A^1_{33} \\ {\scriptstyle 1\times7} \\ C_{23} \end{bmatrix}
\end{bmatrix}}_{\substack{A(\theta) \\ 18\times18}}
\underbrace{\begin{bmatrix} X_{1t-1} \\ X_{2t-1} \\ X_{3t-1} \end{bmatrix}}_{x_{t-1}}
+
\underbrace{\begin{bmatrix}
\underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \vartheta_{gz} & 1 & \cdots & \vdots & 0 \\ \vdots & \cdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}}_{7\times7} \\[6pt]
\begin{bmatrix} B^1_2 \\ {\scriptstyle 4\times3} \end{bmatrix} \begin{bmatrix} 0 \\ {\scriptstyle 4\times4} \end{bmatrix} \\
\begin{bmatrix} B^1_3 \\ {\scriptstyle 1\times7} \\ D_2 \end{bmatrix} \begin{bmatrix} D_2 \end{bmatrix}
\end{bmatrix}}_{\substack{B(\theta) \\ 18\times7}}
\underbrace{\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{et} \\ \varepsilon_{bt} \\ \varepsilon_{ft} \\ \varepsilon_{pt} \\ \varepsilon_{wt} \end{bmatrix}}_{\substack{\varepsilon_t \\ 7\times1}}
$$

$$
\underbrace{\begin{bmatrix} l_t \\ c_t \\ i_t \\ \pi_t \\ w_t \\ r_t \\ y_t \end{bmatrix}}_{\substack{Y_t \\ 7\times1}}
=
\underbrace{\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ {\scriptstyle 1\times7} & {\scriptstyle 1\times7} & {\scriptstyle 1\times7} \\ C_{21} & C_{22} & C_{23} \\ {\scriptstyle 6\times7} & {\scriptstyle 6\times7} & {\scriptstyle 6\times7} \end{bmatrix}}_{\substack{C(\theta) \\ 7\times18}}
\underbrace{\begin{bmatrix} X_{1t-1} \\ X_{2t-1} \\ X_{3t-1} \end{bmatrix}}_{}
+
\underbrace{\begin{bmatrix} D_1 \\ {\scriptstyle 1\times7} \\ D_2 \\ {\scriptstyle 6\times7} \end{bmatrix}}_{\substack{D(\theta) \\ 7\times7}}
\underbrace{\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{gt} \\ \varepsilon_{et} \\ \varepsilon_{bt} \\ \varepsilon_{ft} \\ \varepsilon_{pt} \\ \varepsilon_{wt} \end{bmatrix}}_{}
$$

Figure A.8: Smets and Wouters (2007) ABCD Representation

Note that that the last 5 elements of X_{3t} are the same as the last 5 elements of Y_t, $Y_{1t} = [i_t, \pi_t, w_t, r_t, y_t]'$. In addition, let Y_{0t} be a vector of the first 6 elements of Y_t, i.e. $Y_t = [Y_{0t}', y_t]'$. Finally, define X_{4t} as first 13 elements of X_t, $X_{4t} = [X_{1t}', X_{2t}', l_t, c_t]'$ so

$$\underbrace{\begin{bmatrix} Y_{t+1} \\ \begin{bmatrix} Y_{0t} \\ y_t \end{bmatrix} \\ {}_{Y_t} \end{bmatrix}}_{} = \underbrace{\begin{bmatrix} \underset{13\times13}{F_{11}} & \underset{13\times5}{F_{12}} \\ \underset{1\times13}{F_{21}} & \underset{1\times5}{F_{22}} \end{bmatrix}}_{\begin{bmatrix} CA \\ C \end{bmatrix}} \underbrace{\begin{bmatrix} \underset{13\times1}{X_{4t-1}} \\ \underset{5\times1}{Y_{1t-1}} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} \underset{13\times14}{G_1} \\ \underset{1\times14}{G_2} \end{bmatrix}}_{\begin{bmatrix} D & CB \\ \underset{7\times7}{0} & D \end{bmatrix}} \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \end{bmatrix}$$

The matrix F_{11} is full column rank. Therefore,

$$X_{4t-1} = F_{11}^{-1} \begin{bmatrix} Y_{t+1} \\ Y_{0t} \end{bmatrix} - F_{11}^{-1} F_{12} Y_{1t-1} - F_{11}^{-1} G_1 \begin{bmatrix} \varepsilon_t \\ \varepsilon_t \end{bmatrix}$$

In the An and Schorfheide model, inverting C allowed X_{t-1} to be written as a function of Y_t and ε_t, which ultimately allowed the model to be written in VAR(1) representation. Here, inverting F_{11} allows the 13 states in X_{4t-1} to be written as a function of Y_{t+1}, Y_t, Y_{t-1}, ε_{t+1}, and ε_t, and the remaining 5 states in X_{t-1} besides X_{4t-1} are equal to Y_{1t-1} exactly. First, decompose the matrices A and B as

$$\underset{18\times18}{A} = \begin{bmatrix} \underset{13\times13}{A_{11}^*} & \underset{13\times5}{A_{12}^*} \\ \underset{5\times13}{A_{21}^*} & \underset{5\times5}{A_{22}^*} \end{bmatrix} \qquad \underset{18\times7}{B} = \begin{bmatrix} \underset{13\times7}{B_1^*} \\ \underset{5\times7}{B_2^*} \end{bmatrix}$$

Then, plugging the above expression for X_t into the state equation yields

$$\left(F_{11}^{-1} \begin{bmatrix} Y_t \\ Y_{0t-1} \end{bmatrix} - \begin{bmatrix} \underset{13\times2}{0} & F_{11}^{-1} F_{12} \end{bmatrix} Y_{t-2} - F_{11}^{-1} G_1 \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} \right)$$
$$= A_{11}^* \left(F_{11}^{-1} \begin{bmatrix} Y_{t-1} \\ Y_{0t-2} \end{bmatrix} - \begin{bmatrix} \underset{13\times2}{0} & F_{11}^{-1} F_{12} \end{bmatrix} Y_{t-3} - F_{11}^{-1} G_1 \begin{bmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{bmatrix} \right) + A_{12} Y_{1t-3} + B_1 \varepsilon_{t-2}$$

Second, decompose the following matrices:

$$
\underset{13\times13}{F_{11}A_{11}^*F_{11}^{-1}} =
\left[
\begin{array}{c:c}
\underset{7\times7}{(F_{11}A_{11}^*F_{11}^{-1})_{11}} & \underset{7\times6}{(F_{11}A_{11}^*F_{11}^{-1})_{12}} \\
\hdashline
\underset{6\times7}{(F_{11}A_{11}^*F_{11}^{-1})_{21}} & \underset{6\times6}{(F_{11}A_{11}^*F_{11}^{-1})_{22}}
\end{array}
\right]
\qquad
\underset{13\times5}{F_{11}A_{12}^*} =
\left[
\begin{array}{c}
\underset{7\times5}{(F_{11}A_{12}^*)_1} \\
\hdashline
\underset{6\times5}{(F_{11}A_{12}^*)_2}
\end{array}
\right]
$$

$$
\underset{13\times14}{F_{11}A_{11}^*F_{11}^{-1}G_1} =
\left[
\begin{array}{c:c}
\underset{7\times7}{(F_{11}A_{11}^*F_{11}^{-1}G_1)_{11}} & \underset{7\times7}{(F_{11}A_{11}^*F_{11}^{-1}G_1)_{12}} \\
\hdashline
\underset{7\times7}{(F_{11}A_{11}^*F_{11}^{-1}G_1)_{21}} & \underset{7\times7}{(F_{11}A_{11}^*F_{11}^{-1}G_1)_{22}}
\end{array}
\right]
\qquad
\underset{13\times5}{F_{12}} =
\left[
\begin{array}{c}
\underset{7\times5}{(F_{12})_1} \\
\hdashline
\underset{6\times5}{(F_{12})_2}
\end{array}
\right]
$$

$$
\underset{13\times5}{F_{11}A_{11}^*F_{11}^{-1}F_{12}} =
\left[
\begin{array}{c}
\underset{7\times5}{(F_{11}A_{11}^*F_{11}^{-1}F_{12})_1} \\
\hdashline
\underset{6\times5}{(F_{11}A_{11}^*F_{11}^{-1}F_{12})_2}
\end{array}
\right]
\qquad
\underset{13\times7}{F_{11}B_1^*} =
\left[
\begin{array}{c}
\underset{7\times7}{(F_{11}B_1^*)_1} \\
\hdashline
\underset{6\times7}{(F_{11}B_1^*)_2}
\end{array}
\right]
$$

Multiplying both sides of the previous equation by F_{11} and rearranging ultimately gives the following VARMA(3,2) representation for the observables; note, D is invertible:

$$
Y_t = \underbrace{(F_{11}A_{11}^*F_{11}^{-1})_{11}}_{\substack{\Phi_1(\theta)\\7\times7}}Y_{t-1} + \underbrace{\left(\left[\,(F_{11}A_{11}^*F_{11}^{-1})_{12}\;\vdots\;\underset{7\times1}{0}\,\right] + \left[\,\underset{7\times2}{0}\;\vdots\;(F_{12})_1\,\right]\right)}_{\substack{\Phi_2(\theta)\\7\times7}}Y_{t-2}
$$

$$
+ \underbrace{\left(\left[\,\underset{7\times2}{0}\;\vdots\;(F_{11}A_{12}^*)_1\,\right] - \left[\,\underset{7\times2}{0}\;\vdots\;(F_{11}A_{11}^*F_{11}^{-1}F_{12})_1\,\right]\right)}_{\substack{\Phi_3(\theta)\\7\times7}}Y_{t-3} + \underbrace{D\varepsilon_t}_{U_t}
$$

$$
+ \underbrace{\left(CB - \left(F_{11}A_{11}^*F_{11}^{-1}G_1\right)_{11}\right)D^{-1}}_{\substack{\Delta_1(\theta)\\7\times7}}\underbrace{D\varepsilon_{t-1}}_{U_{t-1}}
$$

$$
+ \underbrace{\left(\left(F_{11}B_1^*\right)_1 - \left(F_{11}A_{11}^*F_{11}^{-1}G_1\right)_{12}\right)D^{-1}}_{\substack{\Delta_2(\theta)\\7\times7}}\underbrace{D\varepsilon_{t-2}}_{U_{t-2}}
$$

Correspondence with Raw Data. The observables Y_t may be decomposed as $Y_t = V_t - V(\theta)$ where V_t is a vector of logged and detrended, but not demeaned, data and V is a logged vector of each data series's unconditional means. For instance, log linearized real interest rates are defined by the linearization above as $r_t = \ln(R_t/R)$, where R is the steady state of R_t. Then, the first element of V_t is

Table A.6: Smets and Wouters model data series

Variable	Series	Notes
L_t	PRS85006023/CE16OV	Seasonally adj., quarterly, end-of-period
C_t	PCE	Seasonally adj., quarterly, end-of-period
I_t	FPI	Seasonally adj., quarterly, end-of-period
P_t	CPIAUCSL	Seasonally adj., quarterly, end-of-period
W_t	COMPNFB	Seasonally adj., quarterly, end-of-period
R_t	1+TB3MS/400	Not S.a., ann. rate, end-of-period, % pts. Math to obtain quarterly gross
Y_t	GDPC1	Seasonally adj., quarterly, end-of-period

$\ln R_t$, logged nominal interest rates, and the first element of V is $\ln(E(R_t))$, where $E(\cdot)$ is the unconditional mean. Where $\phi(\cdot)$ is the function selecting the trend of a given variable and $\Pi_t = P_t/P_{t-1}$ is the gross rate of inflation,

$$Y_t = \begin{bmatrix} l_t \\ c_t \\ i_t \\ \pi_t \\ w_t \\ r_t \\ y_t \end{bmatrix} \equiv V_t - V \qquad V_t = \ln \begin{bmatrix} L_t \\ \frac{C_t}{P_t}/\phi\left(\frac{C_t}{P_t}\right) \\ \frac{I_t}{P_t}/\phi\left(\frac{I_t}{P_t}\right) \\ \Pi_t \\ \frac{W_t}{P_t}/\phi\left(\frac{W_t}{P_t}\right) \\ R_t \\ \frac{Y_t}{P_t}/\phi\left(\frac{Y_t}{P_t}\right) \end{bmatrix} \qquad V = \ln E \begin{bmatrix} L_t \\ \frac{C_t}{P_t}/\phi\left(\frac{C_t}{P_t}\right) \\ \frac{I_t}{P_t}/\phi\left(\frac{I_t}{P_t}\right) \\ \Pi_t \\ \frac{W_t}{P_t}/\phi\left(\frac{W_t}{P_t}\right) \\ R_t \\ \frac{Y_t}{P_t}/\phi\left(\frac{Y_t}{P_t}\right) \end{bmatrix}$$

All data codes below correspond to the St. Louis Federal Reserve's `FRED` database. Empirical V corresponds to the model-theoretic $V(\theta)$ by

$$\ln E \begin{bmatrix} L_t \\ \frac{C_t}{P_t}/\phi\left(\frac{C_t}{P_t}\right) \\ \frac{I_t}{P_t}/\phi\left(\frac{I_t}{P_t}\right) \\ \Pi_t \\ \frac{W_t}{P_t}/\phi\left(\frac{W_t}{P_t}\right) \\ R_t \\ \frac{Y_t}{P_t}/\phi\left(\frac{Y_t}{P_t}\right) \end{bmatrix} \equiv \underbrace{V = V(\theta)}_{\text{mean = steady state}} \equiv \ln \begin{bmatrix} L \\ \widehat{C} \\ \widehat{I} \\ \Pi \\ \widehat{W} \\ R \\ \widehat{Y} \end{bmatrix}$$

Appendix B

Chapter 2 Appendix

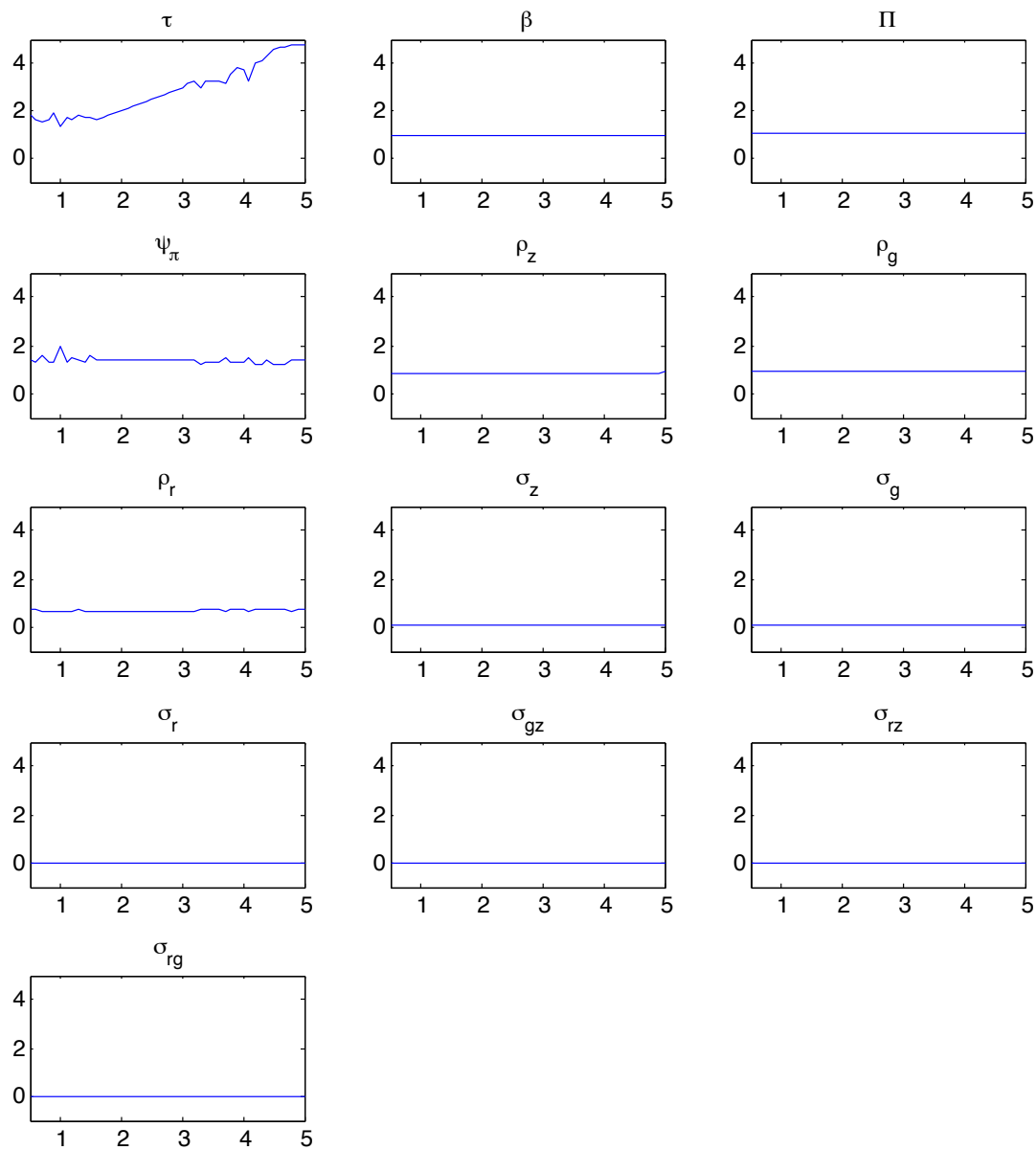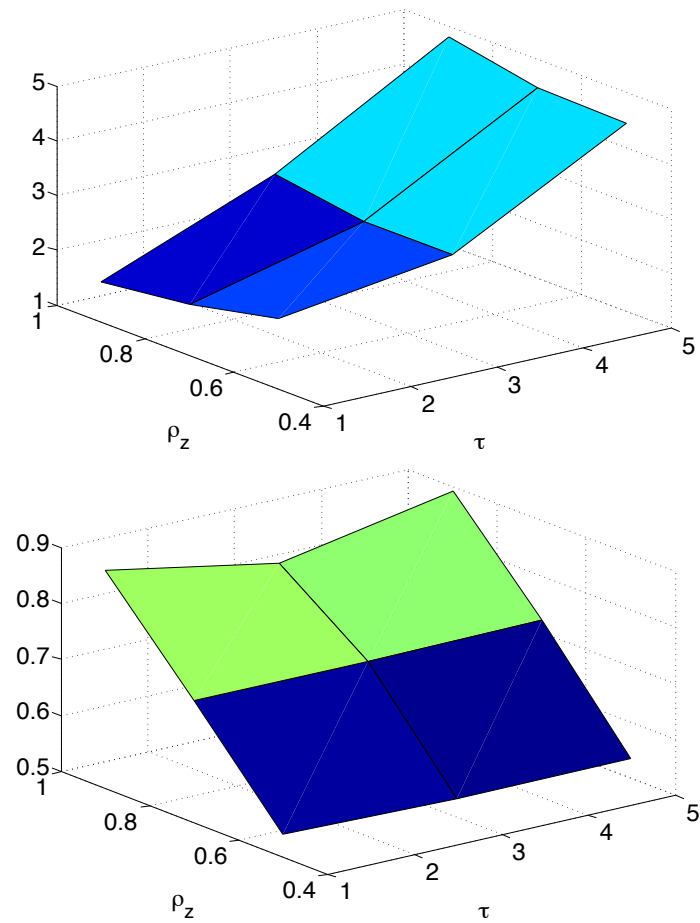**Figure B.1**: Sensitivity of posterior mode (vertical axis) to mean of τ prior (horizontal).

Figure B.2: (Left panel) Sensitivity of τ posterior mode to changes in mean of priors for ρ_z and τ. (Right panel) Same for ρ_z.

Appendix C

Chapter 3 Appendix

A   Proofs

Proof of Proposition 1. See Appendix C.

Proof of Proposition 2. Follows from Kailath et al. (2000) Lemmas E.3.2 and E.4.1 when Proposition 1 is satisfied. See the Appendix to Komunjer and Ng (2011).

Proof of Proposition 3. See Komunjer and Ng (2011).

Proof of Proposition 4. The eigenvalues of P are defined by the characteristic equation $p_P(\lambda) = |P - \lambda I_Z|$. Utilizing the steps in Andreasen et al. (2014)'s proof to their Proposition 1, one may also write $p_P(\lambda) = |h_x - \lambda I_{n_x}||h_x - \lambda I_{n_x}||D_{n_x}^+ h_x^{\otimes 2} D_{n_x} - \lambda I_{n_x(n_x+1)/2}|$. Thus, the eigenvalues of P are determined by $|h_x - \lambda I_{n_x}| = 0$ or $|D_{n_x}^+ h_x^{\otimes 2} D_{n_x} - \lambda I_{n_x(n_x+1)/2}| = 0$. The eigenvalues of the first problem are less than one by Assumption 1, that is $|\lambda_i| < 1$, $i = 1, \ldots, n_x$. To compute the eigenvalues of the second problem, first note that any eigenvalue λ of h_x is determined by the equality $h_x x = \lambda x$ for some eigenvector $x \neq 0$. Therefore, for any two eigenvalues λ_i and λ_j of h_x, $h_x^{\otimes 2}(x_i \otimes x_j) = \lambda_i \lambda_j (x_i \otimes x_j)$. Given the last equality, it must be the case that $h_x^{\otimes 2} D_{n_x} D_{n_x}^+ (x_i \otimes x_j) = \lambda_i \lambda_j (x_i \otimes x_j)$, because $D_{n_x} D_{n_x}^+$ is a matrix which first takes the $n_x(n_x + 1) \times 1$ vech of $x_i x_j'$ using $D_{n_x}^+$, and then duplicates these elements into a new $n_x^2 \times 1$ vector using D_{n_x}. Finally, premultiplying both sides of

the last equality by $D_{n_x}^+$ yields $D_{n_x}^+ h_x^{\otimes 2} D_{n_x} D_{n_x}^+ (x_i \otimes x_j) = \lambda_i \lambda_j D_{n_x}^+ (x_i \otimes x_j)$. Thus, $D_{n_x}^+ (x_i \otimes x_j)$ is an eigenvector of $D_{n_x}^+ h_x^{\otimes 2} D_{n_x}$, which has eigenvalues that are the nonredundant products of h_x's eigenvalues, $\lambda_i \lambda_j$ for $i = 1, \ldots n_x$ and $j = i, \ldots n_x$. Since λ_i and λ_j are less than 1 by Assumption 1, so must $\lambda_i \lambda_j$ for all i and j. $\square$

Proof of Corollary 1. Under Assumptions 1-3, the ABCD model, and hence $\varepsilon_t = \left[u_t' \quad (D_{n_u}(u_t^{\otimes 2} - \text{vec}(I_{n_u^2})))' \quad (n\text{vec}(\widehat{x}_{t-1}^f u_t'))' \right]'$ exists. The covariance matrix of ε_t is written

$$\Sigma_\varepsilon(\theta) = E(\varepsilon_t \varepsilon_t') = \begin{bmatrix} I_{n_u} & E(u_t u_t^{\otimes 2'})D_{n_u}^{+'} & 0_{n_u \times n_n} \\ D_{n_u}^+ E(u_t^{\otimes 2} u_t') & \phi_1 & 0_{n_u(n_u+1)/2 \times n_n} \\ 0_{n_n \times n_u} & 0_{n_n \times n_u(n_u+1)/2} & \phi_2 \end{bmatrix}$$

$$\phi_1 = D_{n_u}^+ E\left[u_t^{\otimes 2} u_t^{\otimes 2'} \right] D_{n_u}^{+'} - D_{n_u}^+ \text{vec}(I_{n_u})\text{vec}(I_{n_u})'D_{n_u}^{+'}$$

$$\phi_2 = nE\left[\text{vec}\left(\widehat{x}_{t-1}^f u_t' \right) \text{vec}\left(\widehat{x}_{t-1}^f u_t' \right)' \right] n'$$

The zeros occur becuase u_t is white noise. The covariance matrix for $\widehat{x}_{t-1}^f$ is defined by the Lyapunov equation $E(\widehat{x}_t^f \widehat{x}_t^{f'}) = h_x E(\widehat{x}_t^f \widehat{x}_t^{f'})h_x' + \sigma^2 \eta \eta'$ and $(I_{n_x^2} - h_x^{\otimes 2})^{-1}$ exists under Assumption 1. Therefore, $E(\widehat{x}_t^f \widehat{x}_t^{f'})$ exists and is finite. In addition given Assumption 3, finite fourth moments for u_t (and hence, also finite third moments), then it follows directly that the entire matrix Σ_ε exists and is finite. However, again because u_t is white noise, all additional moments are

$$E(\varepsilon_t \varepsilon_{t-j}') = \begin{bmatrix} 0_{n_u \times n_u} & 0_{n_u \times n_u(n_u+1)/2} & 0_{n_u \times n_n} \\ 0_{n_u(n_u+1)/2 \times n_u} & 0_{n_u(n_u+1)/2 \times n_u(n_u+1)/2} & 0_{n_u(n_u+1)/2 \times n_n} \\ 0_{n_n \times n_u} & 0_{n_n \times n_u(n_u+1)/2} & 0_{n_n \times n_n} \end{bmatrix} \forall j \geq 1$$

This fact is easy to confirm block-by-block. Thus, ε_t is white noise. $\square$

Proof of Corollary 2. P is related to its eigenvalues and vectors by the relation $Px = \lambda x$ for $x \neq 0$. A is related to P by the relation $A = MPM'$ for M a zero-one selection matrix. Recall, M is constructed to exploit the zero-columns of P so that $P\widehat{Z}_t = PM'M\widehat{Z}_t$. By similar arguments, $MPM'Mx = \lambda Mx$, i.e. $Ay = \lambda y$ for

$y = Mx \neq 0$ an eigenvector of A, and λ, which are A's eigenvalues, are a subset of P's eigenvalues. Proposition 1 guarantees that all $n_x(n_x + 1)/2$ of P's eigenvalues are less than one under Assumption 1. $\square$

Proof of Corollary 3. Utilizing the functional form of A, C, P, and M, the product of C and any power of A has the functional form

$$CA^i = \left[\begin{array}{c|c} 1_2' \otimes (g_x h_x m'(m h_x m')^i) & \cdots \end{array} \right]$$

where $1_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}'$. Therefore, the observability matrix is written

$$\mathcal{O} = \left[\begin{array}{c|c} 1_2' \otimes (g_x h_x m') & \cdots \\ \hline \vdots & \vdots \\ \hline 1_2' \otimes (g_x h_x m'(m h_x m')^{n_X-1}) & \cdots \end{array} \right]$$

Because of the functional form, column j of $\mathcal{O}$ is linearly dependent on column $n_m + j$ for $j = 1, \ldots n_m$ and n_m is the row dimension of m. In other words, the observability matrix has reduced column rank implying the system is not observable. $\square$

B   Pruned State Space Representation

B.1   Baseline Case

The functional form of the matrices in the baseline case state equation second order approximation (3.6) is given by the following steps:

1. H_{xx} is defined as follows: Let $h_{(i)}(x_{t-1}, \sigma|\theta)$ denote row i of $h(x_{t-1}, \sigma|\theta)$. The scalar second-order element of the Taylor expansion of h corresponding to x_t

is

$$\widehat{x}'_{t-1} \times \left. \frac{\partial^2 h_{(i)}(x_{t-1},\sigma|\theta)}{\partial x_{t-1} \partial x'_{t-1}} \right|_{x_{t-1}=x^*,\sigma=0} \times \widehat{x}_t =$$

$$\mathrm{vec}\left(\left. \frac{\partial^2 h_{(i)}(x_{t-1},\sigma|\theta)}{\partial x_{t-1} \partial x'_{t-1}} \right|_{x_{t-1}=x^*,\sigma=0} \right)' \times x_{t-1}^{\otimes 2}$$

and H_{xx} is constructed by stacking the coefficients in rows:

$$\underset{n_x \times n_x^2}{H_{xx}(\theta)} = \begin{bmatrix} \mathrm{vec}\left(\left. \frac{\partial^2 h_{(1)}(x_{t-1},\sigma|\theta)}{\partial x_{t-1} \partial x'_{t-1}} \right|_{x_{t-1}=x^*,\sigma=0} \right)' \\ \vdots \\ \mathrm{vec}\left(\left. \frac{\partial^2 h_{(n_x)}(x_{t-1},\sigma|\theta)}{\partial x_{t-1} \partial x'_{t-1}} \right|_{x_{t-1}=x^*,\sigma=0} \right)' \end{bmatrix}$$

The elements of G_{xx} are defined similarly.

2. The cross-partials between states and σ are

$$\underset{n_x \times n_x}{h_{x\sigma}(\theta)} = \left. \frac{\partial^2 h(x_{t-1},\sigma|\theta)}{\partial \sigma \partial x'_{t-1}} \right|_{x_{t-1}=x^*,\sigma=0} \qquad \underset{n_y \times n_x}{g_{x\sigma}(\theta)} = \left. \frac{\partial^2 g(x_t,\sigma|\theta)}{\partial \sigma \partial x'_t} \right|_{x_t=x^*,\sigma=0}$$

3. $h_{\sigma\sigma}$ and $g_{\sigma\sigma}$ are defined by

$$\underset{n_x \times 1}{h_{\sigma\sigma}(\theta)} = \left. \frac{\partial^2 h(x_{t-1},\sigma|\theta)}{\partial \sigma^2} \right|_{x_{t-1}=x^*,\sigma=0} \qquad \underset{n_x \times 1}{g_{\sigma\sigma}(\theta)} = \left. \frac{\partial^2 g(x_t,\sigma|\theta)}{\partial \sigma^2} \right|_{x_t=x^*,\sigma=0}$$

To obtain the representation (3.13) and (3.14) from the second order pruned state space representation (3.11) and (3.12), observe the equality

$$\widehat{x}^f_t \widehat{x}^{f'}_t = h_x \widehat{x}^f_{t-1} \widehat{x}^{f'}_{t-1} h'_x + \sigma^2 \eta u_t u'_t \eta' + \sigma h_x \widehat{x}^f_{t-1} u'_t \eta' + \sigma \eta u_t \widehat{x}^{f'}_{t-1} h'_x$$

Since $\mathrm{vec}(vv') = v^{\otimes 2}$ for any column vector v,

$$\widehat{x}^{f\otimes 2}_t = \sigma^2 \eta^{\otimes 2} \mathrm{vec}(I_{n_u}) + h_x^{\otimes 2} \widehat{x}^{f\otimes 2}_{t-1} + \sigma(\eta \otimes h_x) \mathrm{vec}(\widehat{x}^f_{t-1} u'_t)$$
$$+ \sigma(h_x \otimes \eta) \mathrm{vec}(u_t \widehat{x}^{f'}_{t-1}) + \sigma^2 \eta^{\otimes 2}(u_t^{\otimes 2} - \mathrm{vec}(I_{n_u})) \quad (\mathrm{C.1})$$

Stacking this underneath Equations (3.9) and (3.11) into a single system immediately yields (3.13) and (3.14).

B.2   Nonlinearity Between Errors and States

The functional form of the matrices in the second order approximation for the state equation allowing for nonlinearities between states and errors (3.16) is given by the following steps:

1. σH_{xu} is defined as follows: Let $h_{(i)}(x_{t-1}, v_t, \sigma|\theta)$ denote row i of $h(x_{t-1}, v_t, \sigma|\theta)$. The scalar second-order element of the Taylor expansion of h corresponding to x_t is

$$
\widehat{x}'_{t-1} \times \left. \frac{\partial^2 h_{(i)}(x_{t-1}, v_t, \sigma|\theta)}{\partial x_{t-1} \partial v'_t} \right|_{x_{t-1}=x^*, u_t=0, \sigma=0} \times (\sigma L_u u_t) =
$$

$$
\sigma \text{vec} \left(\left. \frac{\partial^2 h_{(i)}(x_{t-1}, \sigma|\theta)}{\partial x_{t-1} \partial v'_t} \right|_{x_{t-1}=x^*, u_t=0, \sigma=0} \right)' \times (L_u \otimes I_{n_x}) \times \text{vec}(\widehat{x}_{t-1} u'_t)
$$

and σH_{xu} is constructed by stacking the coefficients in rows:

$$
\underset{n_x \times n_x n_u}{\sigma H_{xu}(\theta)} = \sigma \begin{bmatrix} \text{vec} \left(\left. \frac{\partial^2 h_{(1)}(x_{t-1}, v_t, \sigma|\theta)}{\partial x_{t-1} \partial v'_t} \right|_{x_t=x^*, u_t=0, \sigma=0} \right)' \\ \vdots \\ \text{vec} \left(\left. \frac{\partial^2 h_{(n_x)}(x_{t-1}, v_t \sigma|\theta)}{\partial x_t \partial v'_t} \right|_{x_t=x^*, u_t=0 \sigma=0} \right)' \end{bmatrix} (L_u \otimes I_{n_x})
$$

2. σH_{ux}: Defining $h_{(i)}$ as in Step 1,

$$
(\sigma L_u u_t)' \times \left. \frac{\partial^2 h_{(i)}(x_{t-1}, v_t, \sigma|\theta)}{\partial v_t \partial x'_{t-1}} \right|_{x_{t-1}=x^*, u_t=0, \sigma=0} \times \widehat{x}'_{t-1} =
$$

$$
\sigma \text{vec} \left(\left. \frac{\partial^2 h_{(i)}(x_{t-1}, \sigma|\theta)}{\partial v_t \partial x'_{t-1}} \right|_{x_{t-1}=x^*, u_t=0, \sigma=0} \right)' \times (I_{n_x} \otimes L_u) \times \text{vec}(u_t \widehat{x}'_{t-1})
$$

and σH_{xu} is constructed by stacking the coefficients in rows:

$$\sigma H_{ux}(\theta) = \sigma \begin{bmatrix} \mathrm{vec}\left(\frac{\partial^2 h_{(1)}(x_{t-1},v_t,\sigma|\theta)}{\partial v_t \partial x'_{t-1}}\Big|_{x_{t-1}=x^*,u_t=0,\sigma=0} \right)' \\ \vdots \\ \mathrm{vec}\left(\frac{\partial^2 h_{(n_x)}(x_{t-1},v_t,\sigma|\theta)}{\partial v_t \partial x'_{t-1}}\Big|_{x_{t-1}=x^*,u_t=0,\sigma=0} \right)' \end{bmatrix} (I_{n_x} \otimes L_u)$$

$$\underset{n_x \times n_u n_x}{}$$

3. $\sigma^2 H_{uu}$: Defining $h_{(i)}$ as in Step 1,

$$(\sigma L_u u_t)' \times \frac{\partial^2 h_{(i)}(x_{t-1}, v_t, \sigma|\theta)}{\partial v_t \partial v'_t}\Big|_{x_{t-1}=x^*,u_t=0,\sigma=0} \times (\sigma L_u u_t) =$$

$$\sigma \mathrm{vec}\left(\frac{\partial^2 h_{(i)}(x_{t-1}, \sigma|\theta)}{\partial v_t \partial v'_t}\Big|_{x_{t-1}=x^*,u_t=0,\sigma=0} \right)' \times L_u^{\otimes 2} \times \mathrm{vec}(u_t u'_t)$$

and σH_{uu} is constructed by stacking the coefficients in rows:

$$\sigma H_{xu}(\theta) = \sigma \begin{bmatrix} \mathrm{vec}\left(\frac{\partial^2 h_{(1)}(x_t,\sigma|\theta)}{\partial x_t \partial x'_t}\Big|_{x_t=x^*,\sigma=0} \right)' \\ \vdots \\ \mathrm{vec}\left(\frac{\partial^2 h_{(n_x)}(x_t,\sigma|\theta)}{\partial x_t \partial x'_t}\Big|_{x_t=x^*,\sigma=0} \right)' \end{bmatrix} \times L_u^{\otimes 2}$$

$$\underset{n_x \times n_x n_u}{}$$

B.3 JPRS Representation

Both categories of pruned state space representation described in this paper – the baseline case with linearity in errors and states (3.13) and (3.14), or with nonlinearity in errors and states (3.17) and (3.14) – has the same generalized functional form.

$$\mathcal{Z}_t = \mathcal{J}_{\mathcal{Z}}(\theta) + \mathcal{P}(\theta)\mathcal{Z}_{t-1} + \mathcal{R}(\theta)U_t \tag{C.2}$$

$$\mathcal{Y}_t = \mathcal{J}_{\mathcal{Y}}(\theta) + \mathcal{S}(\theta)\mathcal{Z}_t \tag{C.3}$$

where the variables and matrices of this JPRS representation are defined by

$$
\mathcal{Z}_t = \begin{bmatrix} \widehat{x}_t^f \\ \widehat{x}_t^s \\ \widehat{x}_t^{f \otimes 2} \end{bmatrix} \qquad \mathcal{U}_t = \begin{bmatrix} u_t \\ u_t^{\otimes 2} - \mathrm{vec}(I_{n_u^2}) \\ \mathrm{vec}(\widehat{x}_{t-1}^f u_t') \\ \mathrm{vec}(u_t \widehat{x}_{t-1}^{f'}) \end{bmatrix} \qquad \mathcal{Y}_t = \widehat{y}_t^f + \widehat{y}_t^s
$$

$$
\mathcal{J}_{\mathcal{Z}} = \begin{bmatrix} 0 \\ \frac{1}{2}h_{\sigma\sigma}\sigma^2 + \frac{1}{2}\sigma^2 H_{uu} I_{n_u^2} \\ \sigma^2 \eta^{\otimes 2} \mathrm{vec}(I_{n_u^2}) \end{bmatrix} \qquad \mathcal{J}_{\mathcal{Y}} = \frac{1}{2} g_{\sigma\sigma} \qquad \mathcal{P} = \begin{bmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2}H_{xx} \\ 0 & 0 & h_x^{\otimes 2} \end{bmatrix}
$$

$$
\mathcal{R} = \begin{bmatrix} \sigma\eta & 0 & 0 & 0 \\ 0 & \frac{1}{2}\sigma^2 H_{uu} & \frac{1}{2}\sigma H_{xu} & \frac{1}{2}\sigma H_{ux} \\ 0 & \sigma^2 \eta^{\otimes 2} & \sigma(\eta \otimes h_x) & \sigma(h_x \otimes \eta) \end{bmatrix} \qquad \mathcal{S} = \begin{bmatrix} g_x & g_x & \frac{1}{2}G_{xx} \end{bmatrix}
$$

The dimensions of this representation $n_{\mathcal{Z}} = 2n_x + n_x^2$ and $n_{\mathcal{U}} = n_u + n_u^2 + 2n_x n_u$. Recall, the distinction between the model with nonlinearities between errors and states and the baseline case is that in the latter, H_{uu}, H_{xu}, and H_{ux} are all zero.

C 3-Step ABCD Reparameterization

In order to show that pruned state space representation also has generic ABCD representation, I begin with JPRS representation, given above in Section B .3. Then, I show how to reparameterize the model in three simple steps, the third of which is the most intensive. This section also serves as nontechnical step-by-step proof of Proposition 1.

C.1 Step 1: Remove Redundant States

There are many redundant elements in $\mathcal{Z}_t$ and $\mathcal{U}_t$, given in Section B .3, which we wish to remove. For example, $\widehat{x}_t^f \otimes \widehat{x}_t^f = \mathrm{vec}(x_t^f x_t^{f'})$. Since $x_t^f x_t^{f'}$ is symmetric, $\widehat{x}_t^f \otimes \widehat{x}_t^f$ has exactly $n_x^2 - n_x(n_x+1)/2$ redundant elements, where $n_x(n_x+1)/2$ is the number of elements in $\mathrm{vech}(x_t^f x_t^{f'})$. To handle redundancies such as these, recall that the duplication matrix D_{n_x} is the $n_x \times n_x(n_x + 1)$ dimensional matrix

which yields the equality $\text{vec}(\widehat{x}_t^f \widehat{x}_t^{f'}) = D_{n_x}\text{vech}(\widehat{x}_t^f \widehat{x}_t^{f'})$. $D_{n_x}^+ = (D_{n_x}' D_{n_x})^{-1} D_{n_x}'$ is the Moore-Penrose pseudo inverse of the duplication matrix which yields the equality $D_{n_x}^+ \text{vec}(x_t^f x_t^{f'}) = \text{vech}(x_t^f x_t^{f'})$. In relation, $D_{n_x}^+ D_{n_x} = I_{n_x(n_x+1)/2}$ and $\widehat{x}_t^{f\otimes 2} = D_{n_x} D_{n_x}^+ \widehat{x}_t^{f\otimes 2}$ where $\widehat{x}_t^{f\otimes 2} = \text{vec}(x_t^f x_t^{f'})$. Note, a somewhat nuanced point $D_{n_x} D_{n_x}^+$ is *not* the identity matrix, but a matrix which selects the unique elements of $\widehat{x}_t^{f\otimes 2}$ and uses them to reconstruct the entire vector. Therefore, Equation (3.11) may be rewritten

$$\widehat{x}_t^s \approx h_x(\theta)\widehat{x}_{t-1}^s + \underbrace{\left(\frac{1}{2}H_{xx}(\theta)D_{n_x}\right)}_{n_x \times n_x(n_x+1)} \times \underbrace{\left(D_{n_x}^+ \widehat{x}_{t-1}^{f\otimes 2}\right)}_{\text{vech}(x_{t-1}^f x_{t-1}^{f'})} + \frac{1}{2}h_{\sigma\sigma}(\theta)\sigma^2$$

Meanwhile, D_{n_u} is the $n_u \times n_u(n_u+1)$ dimensional duplication matrix (and $D_{n_u} = (D_{n_u}' D_{n_u})^{-1} D_{n_u}'$ its Moore-Penrose pseudo inverse) which operate on $\text{vec}(u_t u_t')$ (and $\text{vech}(u_t u_t')$) similarly. In addition, K_{n_x,n_u} is the commutation matrix which equates $K_{n_x,n_u}\text{vec}(\widehat{x}_{t-1}^f u_t) = \text{vec}(u_t \widehat{x}_{t-1}^{f'})$. This implies that the product of the first-order solution with itself, Equation (C.1), and also the third block-row of (C.2), may be rewritten

$$D_{n_x}^+ \widehat{x}_t^{f\otimes 2} = \underbrace{\sigma^2 D_{n_x}^+ \eta^{\otimes 2}\text{vec}(I_{n_u})}_{n_x(n_x+1)\times 1} + \underbrace{\left(D_{n_x}^+ h_x^{\otimes 2} D_{n_x}\right)}_{n_x(n_x+1)\times n_x(n_x+1)} \times \left(D_{n_x}^+ \widehat{x}_{t-1}^{f\otimes 2}\right)$$

$$+ \underbrace{\left(\sigma^2 D_{n_x}^+ \eta^{\otimes 2} D_{n_u}\right)}_{n_x(n_x+1)\times n_u(n_u+1)} \times \left(D_{n_u}^+ (u_t^{\otimes 2} - \text{vec}(I_{n_u}))\right) + \underbrace{\sigma r(\theta)}_{n_x(n_x+1)\times n_x n_u} \times \text{vec}(\widehat{x}_{t-1}^f u_t')$$

for $r(\theta) = D_{n_x}^+ \left(\eta \otimes h_x + (h_x \otimes \eta)K_{n_x,n_u}\right)$. Similarly, the second block-row of (C.2) may be written

$$\widehat{x}_t^s = \left(\frac{1}{2}h_{\sigma\sigma}\sigma^2 + \frac{1}{2}\sigma^2 H_{uu}I_{n_u^2}\right) + h_x\widehat{x}_{t-1}^s + \left(\frac{1}{2}H_{xx}D_{n_x}\right) \times (D_{n_x}^+ \widehat{x}_{t-1}^{f\otimes 2})$$

$$+ \left(\frac{1}{2}\sigma^2 H_{uu}D_{n_u}\right) \times (D_{n_u}^+ (u_t^{\otimes 2} - \text{vec}(I_{n_u^2})) + \sigma\frac{1}{2}(H_{xu} + H_{ux}K_{n_x,n_u}) \times \text{vec}(\widehat{x}_{t-1}^f u_t')$$

Together, these equalities imply that the JPRS state in Section B.3 is reducible to

$$\underbrace{\begin{bmatrix} \widehat{x}_t^f \\ \widehat{x}_t^s \\ D_{n_x}^+ \widehat{x}_t^{f\otimes 2} \end{bmatrix}}_{Z_t} = J(\theta) + P(\theta) \times \begin{bmatrix} \widehat{x}_{t-1}^f \\ \widehat{x}_{t-1}^s \\ D_{n_x}^+ \widehat{x}_{t-1}^{f\otimes 2} \end{bmatrix} + R(\theta) \times \underbrace{\begin{bmatrix} u_t \\ D_{n_u}^+ (u_t^{\otimes 2} - \text{vec}(I_{n_u})) \\ \text{vec}(\widehat{x}_{t-1}^f u_t') \end{bmatrix}}_{U_t} \quad \text{(C.4)}$$

where J, P, and R are defined in the text following Proposition 1, equations (3.19) and (3.20) with $n_Z = 2n_x + n_x(n_x+1)/2 < n_{\mathcal{Z}}$ for $n_x > 1$ and $n_U = n_u + n_u(n_u+1)/2 + n_x n_u < n_{\mathcal{U}}$ for all $n_x > 0$ and $n_u > 0$. Similarly, the observation equation is reduced to the following, where $S = \begin{bmatrix} g_x & g_x & \frac{1}{2}G_{xx}D_{n_x} \end{bmatrix}$:

$$\mathcal{Y}_t = \mathcal{K}_{\mathcal{Y}}(\theta) + S(\theta)Z_t \quad \text{(C.5)}$$

C.2 Step 2: Remove Means

It is necessary to remove means from the model to achieve ABCD representation.[1] Before doing so, it is useful to state a simple proposition, related to the findings of Andreasen et al.

Proposition 4. *Under Assumption 1, the eigenvalues of P are less than one.*

A proof of Propostion 4 appears in Appendix A. When it is true, the unconditional mean of the states $E(Z_t|\theta) = E(Z_t|\theta)$ may be defined by

$$E(Z_t|\theta) = (I_{n_Z} - P(\theta))^{-1}J(\theta)$$

Then, the rule of motion for the variables without-means $\widehat{Z}_t = Z_t - E(Z_t|\theta)$ is:

$$\widehat{Z}_t = P(\theta)\widehat{Z}_{t-1} + R(\theta)U_t \quad \text{(C.6)}$$

---

[1]I follow this route with the ultimate intention of applying Komunjer and Ng (2011)'s identification results, which do not take into consideration mean-nonzero models. Iskrev (2010)'s approach to local identification does allow for nonzero means; in order to apply his results, one could alternatively compute the first and second moments of the data set $\{\mathcal{Y}_t\}_{t=1}^T$ from JPRS representation directly using Andreasen et al. (2014)'s approach, and calculate the Jacobian using numerical derivatives. Recall, however, that Iskrev's rank conditions apply to only a finite data sample, whereas Komunjer and Ng's apply to the spectral density.

and the observation equation becomes

$$Y_t = S(\theta)\widehat{Z}_t \tag{C.7}$$

where $Y_t = \mathcal{Y}_t - E(\mathcal{Y}_t|\theta)$ for $E(\mathcal{Y}_t|\theta) = \mathcal{J}_y + SE(Z_t|\theta)$. Recall, I have assumed WLOG that the control variables y_t are the observables. Therefore, Y_t is directly compatible with data that has been separated from its means.

C.3 Step 3: Remove Remaining Unnecessary Variables

Intuition for the Third Step. The final step of the ABCD repararma-terization is the most intensive of the three. In order to motivate it, it is most useful to begin with a simple linear example. Consider a first-order approximation of a simple hypothetical model, with linear dependence between states and shocks, $u_t \sim WN(0,1)$.

$$\underbrace{\begin{bmatrix} \widehat{x}_{1t}^f \\ \widehat{x}_{2t}^f \\ \widehat{x}_{3t}^f \end{bmatrix}}_{\widehat{x}_t^f} = \underbrace{\begin{bmatrix} h_{11} & 0 & h_{13} \\ h_{21} & 0 & h_{23} \\ h_{31} & 0 & h_{33} \end{bmatrix}}_{h_x} \begin{bmatrix} \widehat{x}_{1t-1}^f \\ \widehat{x}_{2t-1}^f \\ \widehat{x}_{3t-1}^f \end{bmatrix} + \sigma \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}}_{\eta} u_t \tag{C.8}$$

The scalar observable will be $\widehat{y}_t^f = g_x \widehat{x}_t^f$. In this model, today's value of $\widehat{x}_{2t}^f$ has no effect on the future values either the states or observables. This is embodied by the exclusion restrictions on h_x that have evidently arisen from the microfoundations of the model. Assumption 2 guarantees that the zeros in h_x do not vary for $\theta \in \Theta$; sometimes, DSGE models with this property are referred to as being "varation-free."

When the zeros in h_x are characteristic of the entire parameter space Θ, one may completely eliminate $\widehat{x}_{2t}^f$ from the solution of the model without loss of generality. First, define the zero-one selection matrix

$$m = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

I call m a "selection" matrix, because because using it to premultiply $\widehat{x}_t^f$, it forms a new vector of the complement of $\widehat{x}_{2t}^f$ in $\widehat{x}_t^f$: $m \times \widehat{x}_t^f = \begin{bmatrix} \widehat{x}_{1t}^f & \widehat{x}_{3t}^f \end{bmatrix}'$. Note, however, that m also has another useful property,

$$m'm \begin{bmatrix} \widehat{x}_{1t-1}^f \\ \widehat{x}_{2t-1}^f \\ \widehat{x}_{3t-1}^f \end{bmatrix} = \begin{bmatrix} \widehat{x}_{1t-1}^f \\ 0 \\ \widehat{x}_{3t-1}^f \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} h_{11} & 0 & h_{13} \\ h_{21} & 0 & h_{23} \\ h_{31} & 0 & h_{33} \end{bmatrix} \begin{bmatrix} \widehat{x}_{1t-1}^f \\ \widehat{x}_{2t-1}^f \\ \widehat{x}_{3t-1}^f \end{bmatrix} = \begin{bmatrix} h_{11} & 0 & h_{13} \\ h_{21} & 0 & h_{23} \\ h_{31} & 0 & h_{33} \end{bmatrix} \begin{bmatrix} \widehat{x}_{1t-1}^f \\ 0 \\ \widehat{x}_{3t-1}^f \end{bmatrix}$$

Thus, with $\widehat{x}_t^f$ and m defined as above, the term $h_x \widehat{x}_t^f$ may simply be replaced with $h_x m' m \widehat{x}_t^f$. Using the above facts, it follows that our model may be exactly rewritten

$$\underbrace{\begin{bmatrix} \widehat{x}_{1t}^f \\ \widehat{x}_{2t}^f \end{bmatrix}}_{X_t = m \widehat{x}_t^f} = \underbrace{\begin{bmatrix} h_{11} & h_{13} \\ h_{31} & h_{33} \end{bmatrix}}_{A} \begin{bmatrix} \widehat{x}_{1t-1}^f \\ \widehat{x}_{2t-1}^f \end{bmatrix} + \underbrace{\begin{bmatrix} \sigma\eta_1 \\ \sigma\eta_2 \end{bmatrix}}_{B} \underbrace{u_t}_{\varepsilon_t}$$

$$\underbrace{\widehat{y}_t^f}_{Y_t} = \underbrace{\begin{bmatrix} g_x h_x^{\cdot 1} & g_x h_x^{\cdot 3} \end{bmatrix}}_{C} \begin{bmatrix} \widehat{x}_{1t-1}^f \\ \widehat{x}_{2t-1}^f \end{bmatrix} + \underbrace{\sigma g_x \eta}_{D} u_t \tag{C.9}$$

where $A = m h_x m'$, $B = \sigma m \eta$, $C = g_x h_x m'$, and D is as expressed above. $h_x^{\cdot i}$ is the entire i-th column of h_x. The dimensions of the state, observables, and innovations are denoted $n_X = 2$, $n_Y = 1$, and $n_\varepsilon = 1$. This is known as the ABCD representation of the model.

As Komunjer and Ng (2011) point out, not only does this process of removing states on the basis of exclusion restrictions reduce the dimension of the state vector for linearized DSGE models, but typically, the remaining state vector $m \widehat{x}_t^f$ is minimal. Since it is easy to use the selection matrix m to obtain minimal ABCD representation of linearized models, a natural question is whether a similar procedure may be used for the class of pruned nonlinear models currently under consideration. This is Step 3.

The Third Step. Consider the pruned second order solution of the same hypothetical model presented in Equation (C.8). Note, due to the nested nature of Taylor approximations, exactly the same h_x is nested in this solution (Compare

with equation (3.11)).

$$
\underbrace{\begin{bmatrix} \widehat{x}^s_{1t} \\ \widehat{x}^s_{2t} \\ \widehat{x}^s_{3t} \end{bmatrix}}_{\widehat{x}^s_t} = \underbrace{\begin{bmatrix} h_{11} & 0 & h_{13} \\ h_{21} & 0 & h_{23} \\ h_{31} & 0 & h_{33} \end{bmatrix}}_{h_x} \begin{bmatrix} \widehat{x}^s_{1t-1} \\ \widehat{x}^s_{2t-1} \\ \widehat{x}^s_{3t-1} \end{bmatrix} + \frac{1}{2} \underbrace{\begin{bmatrix} H_1 & 0 & H_{13} & 0 & 0 & H_{16} \\ H_{21} & 0 & H_{23} & 0 & 0 & H_{26} \\ H_{31} & 0 & H_{33} & 0 & 0 & H_{36} \end{bmatrix}}_{\frac{1}{2}H_{xx}D_{n_x}} \underbrace{\begin{bmatrix} \widehat{x}^{f2}_{1t} \\ \widehat{x}^f_{2t}\widehat{x}^f_{1t} \\ \widehat{x}^f_{3t}\widehat{x}^f_{1t} \\ \widehat{x}^{f2}_{2t} \\ \widehat{x}^f_{3t}\widehat{x}^f_{2t} \\ \widehat{x}^{f2}_{3t} \end{bmatrix}}_{D^+_{n_x}\widehat{x}^f_t} + \frac{1}{2}h_{\sigma\sigma}\sigma^2
$$

Zeros on the first order coefficients often imply zeros on the second order coefficients for the same variable; for intuition, consider the hypothetical process $x_{2t} = \alpha x_{1t-1} + \varepsilon_t$. This explains the location of zeros in H_{xx} in comparison to the zeros in h_x. Returning to Step 2, the representation without-means expression Equation (C.6) for the states is

$$
\underbrace{\begin{bmatrix} \widehat{x}^f_t \\ \widehat{Z}_{2t} \\ \widehat{Z}_{3t} \end{bmatrix}}_{\widehat{Z}_t} = \underbrace{\begin{bmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2}H_{xx}D_{n_x} \\ 0 & 0 & D^+_{n_x}h^{\otimes 2}_x D_{n_x} \end{bmatrix}}_{P} \begin{bmatrix} \widehat{x}^f_{t-1} \\ \widehat{Z}_{2t-1} \\ \widehat{Z}_{3t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \sigma\eta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma^2 D^+_{n_x}\eta^{\otimes 2}D_{n_u} & \sigma r(\theta) \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} u_t \\ u^2_t - 1 \\ \widehat{x}^f_t u_t \end{bmatrix}}_{U_t}
$$

Where $\widehat{Z}_{2t}$ is an $n_x \times 1$ vector of the second order solution states separated from their means, $\widehat{Z}_{2t} = \widehat{x}^s_t - E(\widehat{x}^s_t|\theta_0)$ and $\widehat{Z}_{3t}$ is the $n_x(n_x + 1)/2 \times 1$ mean-zero vector $D^+_{n_x}\widehat{x}^s_t - E(D^+_{n_x}\widehat{x}^s_t|\theta_0)$. The observation equation is the following; $Y_t = y^f_t + y^s_t - \frac{1}{2}g_{\sigma\sigma}\sigma^2 - SE(Z_t|\theta)$ is a scalar:

$$
Y_t = \underbrace{\begin{bmatrix} g_x & g_x & \frac{1}{2}G_{xx}D_{n_x} \end{bmatrix}}_{S} \widehat{Z}_t
$$

The selection matrix m was previously used in the linear case to select a sub-vector of $\widehat{x}^f_t$ corresponding to the non-zero columns of h_x. Within P, however, there are zeros not only in the submatrix h_x, but also the submatrices $\frac{1}{2}H_{xx}D_{n_x}$ and $D^+_{n_x}h^{\otimes 2}_x D_{n_x}$. Thus, defining a similar selection matrix for this case requires a

slightly different strategy. First, drawing on the theme of nestedness of progressively higher-order solutions, note that the only submatrix of P premultiplying $\widehat{Z}_{t-1}$ is again h_x. Thus, with m exactly as previously defined, $m\widehat{Z}_{2t-1}$ is the 3×1 vector that selects only the 3 elements of the 6-dimensional vector $\widehat{Z}_{2t-1}$ that correspond to non-zero columns in h_x. Second, we have the fact that there are zeros in the second, fourth, and fifth columns of $\frac{1}{2}H_{xx}D_{n_x}$ as displayed above. But in addition, note that

$$
D_{n_x}^+ h_x^{\otimes 2} D_{n_x} \atop 6\times 6
=
\left[
\begin{array}{c|c|c|c|c|c}
h_{11}^2 & 0 & h_{11}h_{13} & 0 & 0 & h_{13}^2 \\
\hline
h_{11}h_{21} & 0 & h_{11}h_{23} & 0 & 0 & h_{13}h_{23} \\
\hline
h_{11}h_{31} & 0 & h_{11}h_{33} & 0 & 0 & h_{13}h_{33} \\
\hline
h_{21}^2 & 0 & h_{21}h_{23} & 0 & 0 & h_{23}^2 \\
\hline
h_{21}h_{31} & 0 & h_{21}h_{33} & 0 & 0 & h_{23}h_{33} \\
\hline
h_{31}^2 & 0 & h_{31}h_{33} & 0 & 0 & h_{33}^2
\end{array}
\right]
$$

So, only the second, fourth, and fifth columns of both $\frac{1}{2}H_{xx}D_{n_x}$ and $D_{n_x}^+ h_x^{\otimes 2} D_{n_x}$ have zeros. For this reason define

$$
m^* =
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

Then, $m^*\widehat{Z}_{3t-1}$ is the 3×1 vector that selects only the 3 elements of the 6-dimensional vector $\widehat{Z}_{3t-1}$ that correspond to non-zero columns in $\frac{1}{2}H_{xx}D_{n_x}$ and $D_{n_x}^+ h_x^{\otimes 2} D_{n_x}$, and $m^{*\prime}m\widehat{Z}_{3t-1}$ replaces the appropriate elements with zeros (recall the operations of $m'm$ previously). Thus, constructing the 7×12 matrix M as below, we have

$$
\underbrace{
\begin{bmatrix}
m & 0 & 0 \\
0 & m & 0 \\
0 & 0 & m^*
\end{bmatrix}
}_{M:\ 7\times 12}
\underbrace{
\begin{bmatrix}
\widehat{x}_t^f \\
\widehat{Z}_{2t} \\
\widehat{Z}_{3t}
\end{bmatrix}
}_{\widehat{Z}_t:\ 12\times 1}
=
\underbrace{
\begin{bmatrix}
m\widehat{x}_t^f \\
m\widehat{Z}_{2t} \\
m^*\widehat{Z}_{3t}
\end{bmatrix}
}_{X_t:\ 7\times 1}
$$

Therefore, given the equivalence of representations

$$\widehat{Z}_t = P\widehat{Z}_{t-1} + RU_t \;\;\Leftrightarrow\;\; M\widehat{Z}_t = (MPM')\,M\widehat{Z}_{t-1} + MRU_t$$

$$Y_t = S\widehat{Z}_t \;\;\Leftrightarrow\;\; Y_t = (SPM')\,M\widehat{Z}_{t-1} + SRU_t$$

in terms of $m\widehat{Z}_{2t} = \begin{bmatrix} \widehat{z}_{21t} & \widehat{z}_{23t} \end{bmatrix}'$ and $m^*\widehat{Z}_{3t} = \begin{bmatrix} \widehat{z}_{31t} & \widehat{z}_{33t} & \widehat{z}_{36t} \end{bmatrix}'$ we have

$$\underbrace{\begin{bmatrix} \widehat{x}^f_{1t} \\ \widehat{x}^f_{3t} \\ \widehat{z}_{21t} \\ \widehat{z}_{23t} \\ \widehat{z}_{31t} \\ \widehat{z}_{33t} \\ \widehat{z}_{36t} \end{bmatrix}}_{X_t = M\widehat{Z}_t} = \underbrace{\begin{bmatrix} h_{11} & h_{13} & 0 & 0 & 0 & 0 & 0 \\ h_{31} & h_{33} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_{11} & h_{13} & \frac{1}{2}H_{11} & \frac{1}{2}H_{13} & \frac{1}{2}H_{16} \\ 0 & 0 & h_{13} & h_{33} & \frac{1}{2}H_{31} & \frac{1}{2}H_{33} & \frac{1}{2}H_{36} \\ 0 & 0 & 0 & 0 & h_{11}^2 & h_{11}h_{13} & h_{13}^2 \\ 0 & 0 & 0 & 0 & h_{11}h_{13} & h_{11}h_{33} & h_{13}h_{33} \\ 0 & 0 & 0 & 0 & h_{31}^2 & h_{31}h_{33} & h_{33}^2 \end{bmatrix}}_{A = MPM'} \begin{bmatrix} \widehat{x}^f_{1t-1} \\ \widehat{x}^f_{3t-1} \\ \widehat{z}_{21t-1} \\ \widehat{z}_{23t-1} \\ \widehat{z}_{31t-1} \\ \widehat{z}_{33t-1} \\ \widehat{z}_{36t-1} \end{bmatrix} + MRU_t$$

$$Y_t = \underbrace{\begin{bmatrix} g_x h_x^{\cdot 1} & g_x h_x^{\cdot 3} & g_x h_x^{\cdot 1} & g_x h_x^{\cdot 3} & \chi(1) & \chi(3) & \chi(6) \end{bmatrix}}_{C = SPM'} \begin{bmatrix} \widehat{x}^f_{1t-1} \\ \widehat{x}^f_{3t-1} \\ \widehat{z}_{21t-1} \\ \widehat{z}_{23t-1} \\ \widehat{z}_{31t-1} \\ \widehat{z}_{33t-1} \\ \widehat{z}_{36t-1} \end{bmatrix} + SRU_t$$

where

$$\chi(i) \text{ denotes } g_x \times \left(\text{the } i\text{-th column of } \frac{1}{2}H_{xx}D_{n_x} \right)$$

$$+ \frac{1}{2}G_{xx}D_{n_x} \times \left(\text{the } i\text{-th column of } D^+_{n_x} h_x^{\otimes 2} D_{n_x} \right)$$

There is one last step to reduce the dimension of the errors. Recall, the matrix R is a function of h_{xx} through its submatrix r. Therefore, zeros in h_{xx} will imply elements of U_t may be shed, just as elements of $\widehat{Z}_t$ may be. To see this in the

current ongoing example, and recalling that $r = D_{n_x}^+ (\eta \otimes h_x + (h_x \otimes \eta) K_{n_x, n_u})$, then r has the form

$$
\underset{6 \times 3}{r} =
\begin{bmatrix}
2\eta_1 h_{11} & 0 & 2\eta_1 h_{13} \\
\eta_3 h_{11} + \eta_1 h_{21} & 0 & \eta_3 h_{13} + \eta_1 h_{23} \\
\eta_3 h_{11} + \eta_1 h_{31} & 0 & \eta_3 h_{13} + \eta_1 h_{33} \\
2\eta_3 h_{21} & 0 & 2\eta_3 h_{23} \\
\eta_3 (h_{21} + h_{31}) & 0 & \eta_3 (h_{23} + h_{33}) \\
2\eta_3 h_{31} & 0 & 2\eta_3 h_{33}
\end{bmatrix}
$$

Since r premultiplies $\widehat{x}_{t-1}^f u_t = \begin{bmatrix} \widehat{x}_{1t-1}^f u_t & \widehat{x}_{2t-1}^f u_t & \widehat{x}_{3t-1}^f u_t \end{bmatrix}'$, it is evident how the zeros in h_{xx} corresponding to $\widehat{x}_{2t}^f$ have also translated to zeros in r corresponding to $\widehat{x}_{2t}^f u_t$. Thus, recalling how m was originally constructed, and defining another zero-one matrix

$$
n =
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\rightarrow n'n
\begin{bmatrix}
\widehat{x}_{1t-1}^f u_t \\
\widehat{x}_{2t-1}^f u_t \\
\widehat{x}_{3t-1}^f u_t
\end{bmatrix}
=
\begin{bmatrix}
\widehat{x}_{1t-1}^f u_t \\
0 \\
\widehat{x}_{3t-1}^f u_t
\end{bmatrix}
\quad \text{and} \quad
r
\begin{bmatrix}
\widehat{x}_{1t-1}^f u_t \\
\widehat{x}_{2t-1}^f u_t \\
\widehat{x}_{3t-1}^f u_t
\end{bmatrix}
= r
\begin{bmatrix}
\widehat{x}_{1t-1}^f u_t \\
0 \\
\widehat{x}_{3t-1}^f u_t
\end{bmatrix}
$$

And using n, define N to be

$$
N =
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & n
\end{bmatrix}
$$

$$
\underbrace{\begin{bmatrix} \widehat{x}^f_{1t} \\ \widehat{x}^f_{3t} \\ \widehat{z}_{21t} \\ \widehat{z}_{23t} \\ \widehat{z}_{31t} \\ \widehat{z}_{33t} \\ \widehat{z}_{36t} \end{bmatrix}}_{X_t = M\widehat{Z}_t}
= A \begin{bmatrix} \widehat{x}^f_{1t-1} \\ \widehat{x}^f_{3t-1} \\ \widehat{z}_{21t-1} \\ \widehat{z}_{23t-1} \\ \widehat{z}_{31t-1} \\ \widehat{z}_{33t-1} \\ \widehat{z}_{36t-1} \end{bmatrix}
+ \underbrace{\begin{bmatrix} \sigma\eta_1 & 0 & 0 & 0 \\ \sigma\eta_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \eta_1^2 & 2\eta_1 h_{11} & 2\eta_1 h_{13} \\ 0 & \eta_1\eta_3 & \eta_3 h_{11} + \eta_1 h_{31} & \eta_3 h_{13} + \eta_1 h_{33} \\ 0 & \eta_3^2 & 2\eta_3 h_{31} & 2\eta_3 h_{33} \end{bmatrix}}_{B = MRN'}
\underbrace{\begin{bmatrix} u_t \\ u_t^2 - 1 \\ \widehat{x}^f_{1t} u_t \\ \widehat{x}^f_{3t} u_t \end{bmatrix}}_{\varepsilon_t = NU_t}
$$

$$
Y_t = C \begin{bmatrix} \widehat{x}^f_{1t-1} \\ \widehat{x}^f_{3t-1} \\ \widehat{z}_{21t-1} \\ \widehat{z}_{23t-1} \\ \widehat{z}_{31t-1} \\ \widehat{z}_{33t-1} \\ \widehat{z}_{36t-1} \end{bmatrix}
+ \underbrace{\begin{bmatrix} \sigma g_x \eta & \vdots & \frac{\sigma^2}{2} G_{xx} \eta^{\otimes 2} D_{n_x} & \vdots & \sigma\psi(1) & \vdots & \sigma\psi(3) \end{bmatrix}}_{D = SRN'}
\underbrace{\begin{bmatrix} u_t \\ u_t^2 - 1 \\ \widehat{x}^f_{1t} u_t \\ \widehat{x}^f_{3t} u_t \end{bmatrix}}_{\varepsilon_t = NU_t}
$$

where

$$
\psi(i) = \frac{1}{2} G_{xx} D_{n_x} \times (\text{the } i\text{-th column of } r)
$$

In other words, we now have ABCD representation, with all unnecessary variables in $\widehat{Z}_t$ and U_t eliminated from the system, and Step 3 is completed. In conclusion, under Assumptions 1, 2, and 3, the pruned state space model Equations (3.13) and (3.14) may be written in terms of deviations from mean as an ABCD model

$$
\begin{aligned}
X_t &= A(\theta) \ X_{t-1} + B(\theta) \ \varepsilon_t \\
Y_t &= C(\theta) \ X_{t-1} + D(\theta) \ \varepsilon_t
\end{aligned}
\tag{C.10}
$$

where $X_t = M\widehat{Z}_t$, $\varepsilon_t = NU_t$, $A = MPM'$, $B = MRN'$, $C = SPM'$, and $D = SRN'$ where $\widehat{Z}_t = Z_t - (I_{n_Z} - P)^{-1} K$ for Z_t, U_t, K, P, R, and S defined as in

Equations (C.4) and (C.5) and M and N have the functional form

$$
M = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m^* \end{bmatrix} \qquad
N = \begin{bmatrix} I_{n_u} & 0 & 0 \\ 0 & I_{n_u(n_u+1)/2} & 0 \\ 0 & 0 & n \end{bmatrix}
$$

with the submatrices m, m^*, and n being defined appropriately for the model at hand to eliminate all unnecessary elements of $\widehat{Z}_t$ and U_t. Corollary 1, that ε_t is white noise, in a consequence of the fact that u_t is white noise, and is described in Appendix A.

D Definitions

Definition 1. *Controllability*: *For every $\theta \in \Theta$, define the controllability matrix by*

$$
\mathcal{C}(\theta) = \left(\begin{array}{c|c|c|c} B(\theta) & A(\theta)B(\theta) & \dots & A(\theta)^{n_X-1}B(\theta) \end{array} \right)
$$

I say $\{A(\theta), B(\theta)\}$ is controllable if and only if $\mathcal{C}(\theta)$ is full row rank.

Definition 2. *Observability*: *For every $\theta \in \Theta$, define the observability matrix by*

$$
\mathcal{O}(\theta) = \left(\begin{array}{c|c|c|c} C(\theta)' & A(\theta)'C(\theta)' & \dots & A(\theta)^{n_X-1'}C(\theta)' \end{array} \right)'
$$

I say $\{A(\theta), C(\theta)\}$ is observable if and only if $\mathcal{O}(\theta)$ is full column rank.

Definition 3. *Minimality*: *$\{A(\theta), B(\theta), C(\theta), D(\theta)\}$ is minimal if and only if $\{A(\theta), B(\theta)\}$ is controllable and $\{A(\theta), C(\theta)\}$ is observable (Kailath et al. (2000) p. 765).*

Definition 4. *Stochastic Singularity*: *ABCD representation is stochastically singular ("singular") if $n_\varepsilon \leq n_Y$. If $n_\varepsilon \geq n_Y$, the model is called stochastically nonsingular ("nonsingular"), and if $n_\varepsilon = n_Y$ the model is both singular and nonsingular.*

E   Generalization to Higher Order Models

Andreasen et al. (2014) show that third order pruned state space representation also has JPRS representation. I claim but do not provide formal proof that the above steps may be replicated almost exactly to obtain minimal ABCD or AKCΣ representation of these models. The only additional tool that is required is Meijer (2005)'s triplication and quadruplication matrices used in place of the duplication matrix in Step 1 of the 3-Step ABCD reparameterization. Recall, the duplication matrix has the property of equating $x_t^{\otimes 2}$ to its unique elements only by the equality $x_t^{\otimes 2} = D_{n_x} \times$ (unique elements of $x_t^{\otimes 2}$) (The unique elements of $x_t^{\otimes 2}$ are also vech($x_t x_t'$)). The Moore-Penrose inverse $D_{n_x}^+ = (D_{n_x}' D_{n_x})^{-1} D_{n_x}'$ equates (un. el. of $x_t^{\otimes 2}$) $= D_{n_x}^+ x_t^{\otimes 2}$. The triplication matrix T_{n_x} has the property $x_t^{\otimes 3} = T_{n_x} \times$ (un. el. of $x_t^{\otimes 3}$) and $T_{n_x}^+$ exists. The quadruplication matrix Q_{n_x} has the property that $x_t^{\otimes 4} = Q_{n_x} \times$ (un. el. of $x_t^{\otimes 4}$) and $Q_{n_x}^+$ exists. These matrices and the steps above may be used to obtain minimal representation of third order models. Meijer also provides higher-order n-tuplication matrices that could be used or those interested in fourth or higher order models.

Bibliography

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.

ANDREASEN, M. M., J. FERNÁNDEZ-VILLAVERDE, AND J. RUBIO-RAMÍREZ (2014): "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications," *NBER Working Paper Series*, Working Paper 18983.

ANDREWS, I. AND A. MIKUSHEVA (2013): "Maximum Likelihood Inference in Weakly Identified DSGE Models," *Under Review, Quantitative Economics*.

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

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

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

CHARI, V., P. J. KEHOE, AND E. R. MCGRATTAN (2009): "New Keynesian Models: Not Yet Useful for Policy Analysis," *American Economic Journal: Macroeconomics*, 1, 242–266.

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

CHRISTIANO, L. J., M. TRABANDT, AND K. WALENTIN (2010): "DSGE Models for Monetary Policy Analysis," *NBER Working Paper No. 16074*.

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

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

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

FERNÁNDEZ-VILLAVERDE, J., J. RUBIO-RAMÍREZ, T. J. SARGENT, AND M. W. WATSON (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021–1026.

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

FUKAČ, M., D. F. WAGGONER, AND T. ZHA (2007): "Local and Global Identification of DSGE Models: A Simultaneous-Equation Approach," *Working Paper*.

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

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

GALÍ, J., F. SMETS, AND R. WOUTERS (2011): "Unemployment in and Estiamted New Keynesian Model," in *Unemployment in an Estimated New Keynesian Model*, NBER Macroeconomics Annual, 329–60.

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

GUERRON-QUINTANA, P., A. INOUE, AND L. KILIAN (2013): "Frequentist Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 4, 197–229.

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

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

HAMILTON, J. D., D. WAGGONER, AND T. ZHA (2007): "Normalization in Econometrics," *Econometric Reviews*, 26, 221–252.

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

HANSEN, L. P. AND T. SARGENT (2005): *Recursive Methods of Dynamic Linear Economies*, Unpublished Manuscript, Unviersity of Chicago.

HERBST, E. AND F. SCHORFHEIDE (2014): "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, Forthcoming.

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

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

KAILATH, T., A. H. SAYED, AND B. HASSIBI (2000): *Linear Estimation*, Prentice Hall.

KIM, J., S. KIM, E. SCHAUMBERG, AND C. SIMS (2008): "Calculating and Using Second-Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models," *Journal of Economic Dynamics and Control*, 32, 3397–3414.

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

KOCIĘCKI, A. AND M. KOLASA (2013): "Global Identificaiton of Linearized DSGE Models," *Working Paper*.

KOMUNJER, I. AND S. NG (2011): "Dynamic Identification of DSGE Models," *Econometrica*, 79, 1995–2032.

KOOP, G., M. H. PESARAN, AND R. P. SMITH (2013): "On Identification of Bayesian DSGE Models," *Journal of Business and Economic Statistics*, 31, 1995–2032.

KOOPMANS, T. C. AND O. REIERSØL (1950): "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21.

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

LOMBARDO, G. AND A. SUTHERLAND (2007): "Computing second-order-accurate solutions for rational expectations models using linear solution methods," *Journal of Economic Dynamics & Control*, 31, 515–30.

LUTKEPOHL, H. (2005): *New Introduction to Multiple Time Series Analysis*, Springer.

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

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

MEIJER, E. (2005): "Matrix Algebra for Higher Order Moments," *Linear Algebra And Its Applications*, 410, 112–134.

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

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

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

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

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

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

RÍOS-RULL, J.-V., F. SCHORFHEIDE, C. FUENTES-ALBERO, M. KRYSHKO, AND R. SANTAEULÀLIA-LLOPIS (2012): "Methods Versus Substance: Measuring the Effects of Technology Shocks on Hours," *Journal of Monetary Economics*, 59, 826–846.

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

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

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

SCHMITT-GROHÉ, S. AND M. URIBE (2004): "Solving dynamic general equilibrium models using a second-order approximation to the policy function," *Journal of Economic Dynamics and Control*, 28, 755–775.

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.

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

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

VAN BINSBERGEN, J. H., J. FÉRNANEDEZ-VILLAVERDE, R. S. J. KOIJEN, AND J. F. RUBIO-RAMÍREZ (2012): "The Term Structure of Interest Rates in a DSGE Model with Recursive Preferences," *Journal of Monetary Economics*, 59.

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

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