Local and Global Parameter Identification in DSGE Models Allowing for Indeterminacy*

Zhongjun Qu† Boston University

Denis Tkachenko‡ National University of Singapore

December 19, 2012; this version: July 31, 2014

Abstract

This paper presents a unified framework for analyzing local and global identification in log linearized DSGE models that encompasses both determinacy and indeterminacy. The analysis is conducted from a frequency domain perspective. First, for local identification, it presents necessary and sufficient conditions for: (1) the identification of the structural parameters along with the sunspot parameters, (2) the identification of the former irrespective of the latter and (3) the identification of the former conditional on the latter. These conditions apply to both singular and nonsingular models and also permit checking whether a subset of frequencies can deliver identification. Second, for global identification, the paper considers a frequency domain expression for the Kullback-Leibler distance between two DSGE models and shows that global identification fails if and only if the minimized distance equals zero. As a by-product, it delivers parameter values that yield observational equivalence under identification failure. This condition requires nonsingularity but can be applied to nonsingular subsystems and across models with different structures. Third, to develop a further understanding of the strength of identification, the paper proposes a measure for the empirical closeness between two DSGE models. The measure gauges the feasibility of distinguishing one model from another using likelihood ratio tests based on a finite number of observations generated by the two models. The theory is illustrated using two small scale and one medium scale DSGE models. The results document that parameters can be identified under indeterminacy but not determinacy, that different monetary policy rules can be (nearly) observationally equivalent, and that identification properties can differ substantially between small and medium scale models.

Keywords: Dynamic stochastic general equilibrium models, frequency domain, global identification, multiple equilibria, spectral density.

JEL Classification: C10, C30, C52, E1,E3.

*We are grateful to Bertille Antoine, Jesús Fernández-Villaverde, Frank Kleibergen, Arthur Lewbel, Eric Renault, Yixiao Sun and seminar participants at Academia Sinica, BC, Brown, MSU, Birmingham Macroeconomics and Econometrics Conference, CIREQ Econometrics Conference, Econometric Society Winter Meeting and NBER summer institute for useful comments and suggestions.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu).

‡Department of Economics, National University of Singapore (ecstd@nus.edu.sg).

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models provide a unified framework for analyzing business cycles, understanding monetary policy and forecasting. An inherent feature of such models is that, under some parameter values, there can exist a continuum of stable solutions, all satisfying the same set of conditions for equilibrium. This feature, called indeterminacy, can be valuable for understanding the sources and dynamics of macroeconomic fluctuations. For example, Lubik and Schorfheide (2004) considered a prototypical DSGE model and argued that indeterminacy is consistent with the way US monetary policy was conducted over the period 1960:I to 1979:II. Related work in this area includes Leeper (1991), Clarida, Galí and Gertler (2000), Benhabib, Schmitt-Grohé and Uribe (2001), Boivin and Giannoni (2006), Benati and Surico (2009), Mavroeidis (2010) and Cochrane (2011, 2014). Benhabib and Farmer (1999) documented various economic mechanisms leading to indeterminacy, suggesting a further integration of this feature into the DSGE theory.

Parameter identification in DSGE models is important for both calibration and formal statistical analysis. Substantial progress has been made recently. Canova and Sala (2009) documented the types of identification issues that can arise when analyzing DSGE models. Iskrev (2010) gave sufficient conditions, while Komunjer and Ng (2011) and Qu and Tkachenko (2012) provided necessary and sufficient conditions for local identification. However, these conditions have two substantial limitations. First, they assume determinacy. Second, they are silent about global identification. Furthermore, the DSGE literature contains no published work that provides necessary and sufficient conditions for global identification, even under determinacy, except when the solution can be presented as a finite order vector autoregression (Rubio-Ramírez, Waggoner and Zha, 2010 and Fukač, Waggoner and Zha, 2007). This paper aims to make progress in three directions. First, it presents necessary and sufficient conditions for local identification that encompass both determinacy and indeterminacy. Second, it provides a necessary and sufficient condition for global identification. Third, it proposes a measure to gauge the empirical closeness of DSGE models, reflecting the feasibility of distinguishing one model from another based on finite sample sizes.

The local identification analysis builds on the work of Lubik and Schorfheide (2003) and extends the results in Qu and Tkachenko (2012). It proceeds in three steps. First, it defines an augmented parameter vector that consists of both the structural (i.e., those appearing in the log linearized structural model) and the sunspot (i.e., those appearing only under indeterminacy) parameters. Next, it constructs a spectral density which characterizes the second order properties of the full set

of solutions. Third, it treats this spectral density as an infinite dimensional mapping, as in Qu and Tkachenko (2012), studying its properties under local perturbation of parameter values to obtain characterizing conditions for identification. The analysis exploits two features of DSGE models. That is, the models are general equilibrium models and the solutions are vector linear processes. The first feature makes the issue of observed exogenous processes irrelevant. Together, the two features imply that the spectral density summarizes all relevant information under normality, which holds under both determinacy and indeterminacy.

The results on local identification cover three situations. (1) The identification of both the structural and the sunspot parameters. This reveals the possibility of locally determining both the parameters describing technology and preferences and those governing the equilibrium beliefs of agents. (2) The identification of the structural parameters without making statements about the identification of the sunspot parameters. This shows whether it is possible to locally determine the parameters describing technology and preferences, even though the equilibrium beliefs may be unidentifiable. (3) The identification of the structural parameters conditional on the value of the sunspot parameters. This reveals whether it is possible to locally determine the parameters describing technology and preferences once we select a mechanism for equilibrium belief formation. In each of the three situations, the paper provides necessary and sufficient conditions for identification. In practice, the conditions can be applied sequentially. If the first condition is violated, then the other two can be used to better understand the sources of identification failure. All the results allow arbitrary relationships between the number of observables and shocks, therefore are applicable to both singular and nonsingular DSGE models. They permit latent state variables and measurement errors and are relatively straightforward to compute.

The global identification analysis in DSGE models faces two challenges. (1) The model's solutions are often not representable as regressions with a finite number of predetermined regressors, especially under indeterminacy. (2) The parameters interact in a highly nonlinear fashion. These two features make the results in Rothenberg (1971, Sections 4-6) inapplicable. Exploring a different route, this paper obtains the identification condition by making further uses of the general equilibrium feature and vector linear structure of the solutions. First, it introduces a frequency domain expression for the Kullback–Leibler distance between two DSGE models. This criterion depends only on the spectral densities and the dimension of the observables, thus can be computed without simulation or any reference to any data. Second, it shows that global identification fails if and only if the criterion equals zero when minimized over the relevant region of the parameter space. When

identification fails, the condition delivers a set of parameter values that lead to observational equivalence. Meanwhile, when global identification holds, the minimized value is still informative about the strength of the identification. The latter feature is further exploited in the paper to develop a measure for the empirical closeness between models. In relation to the literature, Fernández-Villaverde and Rubio-Ramírez (2004) are among the first to consider the Kullback–Leibler distance in the context of DSGE models. They showed that the parameter estimates obtained from Bayesian methods converge to their pseudo true values and the best model under the Kullback–Leibler distance will receive the highest posterior probability. The current paper is the first that makes use of the Kullback–Leibler distance for checking global identification. The feasibility lies in the feature that its frequency domain expression only depends on the spectral densities, which are easily computable even under indeterminacy, as illustrated later in the paper.

The proposed local and global identification conditions can be compared along three dimensions. First and broadly, identification analysis can involve two types of questions with increasing levels of complexity, i.e., parameter identification conditional on the model structure and identification permitting different structures. In the current context, the former asks whether there exists a different parameter value within the same DSGE structure that can lead to the same dynamics for the observables. The latter asks whether an alternatively specified DSGE structure (for example, a model with a different monetary policy rule or with different shock processes) can generate the same dynamic properties as some benchmark structure. The local identification condition, here and elsewhere in the literature, can only be used to address the first type of identification question. In contrast, the global condition proposed here can also address the second type. This is the only condition in the DSGE literature with such a property. Second, the computational cost associated with the global identification condition can be substantially higher than in the case of the local condition when the dimension of the parameter vector is high. Nevertheless, this paper provides ample evidence showing that it can be effectively implemented for both small scale and medium scale DSGE models. Third, the global identification condition requires the model to be nonsingular. For singular systems, it can still be applied to nonsingular subsystems, as shown in the paper.

When global identification holds, there may still exist parameter values that are difficult to distinguish based on finite samples. For example, Del Negro and Schorfheide (2008) considered a New Keynesian DSGE model and observed that the data provides similar support for a model with moderate price and trivial wage rigidity and one in which both rigidities are high. In Section 7.2, this paper finds that apparently very different Taylor rule parameters can lead to near observational

equivalence in a model considered by Lubik and Schorfheide (2004). More generally, even models with different structures (i.e., with different policy rules or different shock processes) can be similar quantitatively. To address this issue, this paper develops a measure for the empirical closeness between DSGE models. The measure gauges the feasibility of distinguishing one model from another using likelihood ratio tests based on a finite number of observations generated by the two models. It has three features. First, it is straightforward to compute for general DSGE models. The main computation cost is in solving the two models once and computing their respective spectral densities. Second, its value always falls between 0 and 1 with a higher value indicating a greater difference. Third, it monotonically approaches one as the sample size increases if the Kullback–Leibler criterion is positive. The development here is related to Hansen (2007) who, working from a Bayesian decision making perspective, proposed to use Chernoff's (1952) information measure to quantify statistical challenges for distinguishing between two models.

The above methods are applied to three DSGE models. The first two are small scale models considered in An and Schorfheide (2007) and Lubik and Schorfheide (2004), while the third is the medium scale model of Smets and Wouters (2007). Putting these three models together showcases how identification properties can differ within small scale models and between small and medium scale models. The main findings can be summarized as follows. First, consider the model of An and Schorfheide (2007). Previously, Qu and Tkachenko (2012) showed that the Taylor rule parameters are locally unidentified at a parameter value that yields determinacy. Here, the paper considers a range of different values under indeterminacy and finds consistently that the Taylor rule parameters are globally identified. Such a distinction in parameter identification has previously only been documented for models with analytical solutions. The result here shows that this can occur in models of empirical relevance. When measuring the empirical distance between models with different monetary policy rules, the method detects an expected inflation rule that leads to exact observational equivalence with the current inflation rule. Meanwhile, there exists an output growth rule that leads to near observational equivalence as the output gap rule. Both findings hold under both determinacy and indeterminacy. Next, consider the model analyzed in Lubik and Schorfheide (2004). In contrast to the previous model, the paper finds that the parameters are globally identified at the posterior mean under both determinacy and indeterminacy. Meanwhile, the exact and near observational equivalence between alternative monetary policy rules still persists. Third, consider the model of Smets and Wouters (2007). The model is globally identified at the posterior mean under determinacy after fixing five parameters as in the original paper. However,

some parameters (i.e., the steady state elasticity of the capital adjustment cost function and the elasticity of labor supply) are identified only weakly. In contrast to the small scale models, the exact and near observational equivalence between monetary policy rules are no longer present. When measuring the empirical distance between models with different degrees of real and nominal frictions, the results suggest that all eight frictions are empirically relevant in generating the model's dynamic properties. The least important frictions are, among the nominal frictions, the price and wage indexation and, among the real frictions, the elasticity of capital utilization adjustment cost. Interestingly, the conclusions reached regarding both the nominal and real frictions are consistent with those obtained from a Bayesian perspective in Smets and Wouters (2007).

This paper expands the early literature on identification in rational expectations models. There, notable contributions include Wallis (1980), Pesaran (1981) and Blanchard (1982). Wallis (1980, p.63) provided a rank condition for the local identification of the structural parameters in a model that features observable exogenous processes and lagged expectations (i.e., the structural equations at time t include only conditional expectations formed at $t-1$). These two features make the model effectively static. Pesaran (1981, Theorem 1) considered the same model and provided a necessary and sufficient condition for global identification of the structural parameters in a particular structural equation. He also considered models under indeterminacy. However, no formal identification conditions were provided for this case. Blanchard (1982) extended Wallis' (1980) analysis to first-order dynamic linear models. The above analyses share a common feature. That is, the solutions are finite order static regressions or vector autoregressions. This makes identification conditions for standard simultaneous equations system applicable. Therefore, although their results shed some light on the problem, their methods are not applicable to general DSGE models. This paper also contributes to the literature that studies dynamic equilibrium models from a frequency domain perspective. Related studies include Altug (1989), Christiano, Eichenbaum and Marshall (1991), Sims (1993), Hansen and Sargent (1993), Watson (1993), Diebold, Ohanian and Berkowitz (1998), Christiano and Vigfusson (2003) and Del Negro, Diebold and Schorfheide (2008). The current paper is the first to study global identification in DSGE models from the frequency domain perspective.

The paper is structured as follows. Section 2 discusses the model's solutions and the spectral density under indeterminacy. Sections 3 and 4 present results on local and global identification, respectively. Section 5 proposes an empirical distance measure. Section 6 discusses the application of the results to (factor augmented) VARMA models. Section 7 provides some illustrative applications while Section 8 concludes. The paper has three appendices. Appendix A contains details

on solving the model under indeterminacy following Lubik and Schorfheide (2003). Appendix B outlines the model of Smets and Wouters (2007). All proofs are included in Appendix C.

The following notation is used. i is the imaginary unit. $||x||$ is the Euclidean norm of a vector x. $||X||$ is the vector induced norm for a complex valued matrix X. X^* stands for its conjugate transpose while X' is the transpose without conjugation. If $f_\theta \in R^k$ is a differentiable function of $\theta \in R^p$, then $\partial f_{\theta_0}/\partial\theta'$ is a k-by-p matrix of partial derivatives evaluated at θ_0. "$\rightarrow^d$" signifies convergence in distribution and $O(\cdot)$ and $o(\cdot)$ are the usual symbols for orders of magnitude.

2 The spectrum of a DSGE model under indeterminacy

Suppose a DSGE model has been log linearized around its steady state (Sims, 2002):

$$\Gamma_0 S_t = \Gamma_1 S_{t-1} + \Psi\varepsilon_t + \Pi\eta_t, \tag{1}$$

where S_t is a vector that includes the endogenous variables, the conditional expectations and variables from exogenous shock processes if they are serially correlated. The vector ε_t contains serially uncorrelated structural shocks and η_t contains expectation errors. The elements of Γ_0, Γ_1, Ψ and Π are functions of the structural parameters in the model. Depending on the properties of Γ_0 and Γ_1, the system can have none, a unique or multiple stable solutions. This paper assumes at least one stable solution exists and focuses on such cases. For local identification analysis, the relationship between the number of endogenous variables and structural shocks is allowed to be arbitrary, that is, the system can be singular.

Under indeterminacy, the structural parameters alone do not uniquely determine the dynamics (i.e., the spectral density or the autocovariances) of the observables. This constitutes a conceptual challenge for analyzing identification. We take three key steps to overcome this obstacle. First, the result in Lubik and Schorfheide (2003) is applied to obtain a representation for the full set of solutions under indeterminacy. We also amend their procedure to allow the degree of indeterminacy to exceed one. Second, the parameter space is augmented to include the sunspot parameters in addition to the structural parameters. Third, utilizing the augmented parameter space, we compute the unique spectral density which fully characterizes the second order properties of the observables. This spectral density is easily computable and serves as the main instrument for the identification analysis in this paper. The details of the three steps are as follows.

Step 1. Model solution under indeterminacy. Lubik and Schorfheide (2003) showed that under indeterminacy, the full set of solutions can be represented as

$$S_t = \Theta_1 S_{t-1} + \Theta_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t, \tag{2}$$

or equivalently,

$$S_t = (1 - \Theta_1 L)^{-1} [\Theta_\varepsilon \;\; \Theta_\epsilon] \begin{pmatrix} \varepsilon_t \\ \epsilon_t \end{pmatrix},$$

where L denotes the lag operator. Appendix A contains an alternative derivation of this representation using an elementary result in matrix algebra (see the discussion between (A.5) and (A.7)). There, the reduced column echelon form is used to represent the indeterminacy space when the dimension of the latter exceeds one. In the above, the matrices Θ_1, Θ_ε and Θ_ϵ depend only on Γ_0, Γ_1, Ψ and Π, therefore are functions of the structural parameters only. The dimension of Θ_ϵ, called the degree of indeterminacy, is also determined by the structural parameters. As in the literature, the vector ϵ_t is labeled as sunspot shocks.

It is important to emphasize that the DSGE model alone imposes little restriction on ϵ_t. It needs to be a martingale difference, i.e., $E_t \epsilon_{t+1} = 0$, however can be arbitrarily contemporaneously correlated with the fundamental shocks ε_t. Intuitively, the properties of ϵ_t depend on how agents form their expectations, which, under indeterminacy, is not fully revealed by the structural model itself. In light of this, we write

$$\epsilon_t = M \varepsilon_t + \widetilde{\epsilon}_t$$

and make the following assumption, allowing ϵ_t to be contemporaneously correlated with ε_t, being conditional heteroskedastic (e.g., as in GARCH type models) with an arbitrary distribution.

Assumption 1 *$E\left(\widetilde{\epsilon}_t \widetilde{\epsilon}_t'\right) = \Sigma_\epsilon$ with $0 < ||\Sigma_\epsilon|| < \infty$, $||M|| < \infty$, $E\left(\varepsilon_t \widetilde{\epsilon}_s'\right) = 0$ for all t,s and $E\left(\widetilde{\epsilon}_t \widetilde{\epsilon}_s'\right) = 0$ for all $t \neq s$.*

Step 2. Parameter augmentation. Let θ^D be a p-by-1 vector consisting of all the structural parameters in (1). Let θ^U be a q-by-1 vector consisting of the sunspot parameters:

$$\theta^U = \begin{pmatrix} \text{vec}\left(\Sigma_\epsilon\right) \\ \text{vec}\left(M\right) \end{pmatrix}.$$

Define an augmented parameter vector:

$$\theta = \begin{pmatrix} \theta^D \\ \theta^U \end{pmatrix}.$$

Step 3. Computing the spectral density under indeterminacy. In practice, the estimation or calibration is typically based on a subset of S_t or some linear transformations involving the current and lagged values of S_t. To match such a practice, we let $A(L)$ denote a matrix of finite order lag polynomials to specify the observables and write

$$Y_t(\theta) = A(L)S_t = H(L;\theta)\begin{pmatrix} \varepsilon_t \\ \epsilon_t \end{pmatrix}, \tag{3}$$

where

$$H(L;\theta) = A(L)(1-\Theta_1 L)^{-1}[\Theta_\varepsilon \;\; \Theta_\epsilon]. \tag{4}$$

In the above notation, the observables are explicitly indexed by θ to signify that their dynamic properties will depend on both the structural and the sunspot parameters. The matrix $A(L)$ permits analyzing models with latent endogenous variables such as Smets and Wouters (2003, 2007). In these two models, S_t includes variables from a frictionless economy unobservable to the econometrician. Such variables can be excluded simply by assigning zero entries in $A(L)$.

Remark 1 *The representation given by (3) and (4) demonstrates that $Y_t(\theta)$ is a vector linear process driven by ε_t and ϵ_t and that it admits a vector moving average representation irrespective of the number of shocks in the model. It often does not have a finite order VAR representation. For example, consider the leading case with a nonsingular system where the dimension of ε_t equals that of $Y_t(\theta)$. Then, because of the presence of ϵ_t, the system is non-invertible, precluding any vector autoregressive representations in the first place. The polynomial $H(L;\theta)$ can be of infinite order due to the presence of $A(L)$ and $(1-\Theta_1 L)^{-1}$. This implies that the solutions can possess an infinite number of reduced form parameters, making the latter unidentifiable without additional restrictions.*

The spectral density of $Y_t(\theta)$ is, using (3) and (4),

$$f_\theta(\omega) = \frac{1}{2\pi}H(\exp(-i\omega);\theta)\Sigma(\theta)H(\exp(-i\omega);\theta)^*, \tag{5}$$

where

$$\Sigma(\theta) = \begin{pmatrix} I & 0 \\ M & I \end{pmatrix}\begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix}\begin{pmatrix} I & 0 \\ M & I \end{pmatrix}'.$$

Note that adding measurement errors would present no difficulty: suppose we observe $Y_t(\theta)+\zeta_t(\theta)$, with $\zeta_t(\theta)$ being measurement errors independent of $Y_t(\theta)$ with spectrum $f^m_\theta(\omega)$, then the spectral density of the observables is given by

$$f_\theta(\omega) = \frac{1}{2\pi}H(\exp(-i\omega);\theta)\Sigma(\theta)H(\exp(-i\omega);\theta)^* + f^m_\theta(\omega),$$

where $H(\cdot)$ is again defined by (4). The dimension of $f_\theta(\omega)$ is always n_Y-by-n_Y with n_Y being the dimension of $Y_t(\theta)$. In particular, its dimension does not depend on the number of shocks or whether there is indeterminacy.

Throughout the paper, we let $\{Y_t\}$ denote a stochastic process whose spectral density is given by $f_{\theta_0}(\omega)$ for $\omega \in [-\pi, \pi]$. The next assumption imposes some structure on the parameter space and the spectral density.

Assumption 2 *(i) Assume $\theta^D \in \Theta^D \subset \mathbb{R}^p$ and $\theta \in \Theta \subset \mathbb{R}^{p+q}$ with Θ^D and Θ being compact. Assume θ_0^D is an interior point of Θ^D and θ_0 is an interior point of Θ under indeterminacy. (ii) The elements of the spectral density $f_\theta(\omega)$ are continuous in ω and continuous and differentiable in θ; $\|f_\theta(\omega)\| \leq C < \infty$ for all $\theta \in \Theta$ and $\omega \in [-\pi, \pi]$.*

For local identification, the compactness is not required. Also, (ii) only needs to hold in an open neighborhood of θ_0 over $\omega \in [-\pi, \pi]$. If unit roots are present in the DSGE model, then the assumption requires appropriately transforming the series prior to applying the methods.

Remark 2 *Influenced by Koopmans (1949, Abstract p. 125), the identification analysis in econometrics is often embedded in a hierarchical or two step formulation. First, the joint distribution of the observations is written in terms of reduced form equations in which the parameters are always identifiable. Second, the structural parameters are linked to the reduced form parameters through a finite number of time invariant restrictions, which uniquely matters for the identification. However, as discussed in Remark 1, in DSGE models the reduced form often does not possess a natural definition and the parameters involved can be unidentifiable. This motivates us to adopt Haavelmo's (1944, p. 99) formulation to write the joint distribution (more precisely, the spectral density) directly in terms of the structural parameters. This basic principle underlies the analysis of both local and global identification in this paper.*

3 Local identification

Komunjer and Ng (2011) and Qu and Tkachenko (2012) discussed potential challenges for studying local identification in DSGE models under determinacy. Similar challenges persist under indeterminacy. First, as stated in Remarks 1 and 2, $Y_t(\theta)$ in (3) in general does not admit a finite order vector autoregressive representation. This implies that the conventional approach of first locating the identified finite dimensional reduced form parameters and then using them to back out the structural ones breaks down. Second, because this is a general equilibrium model, it typically does

not include any observed exogenous processes. Therefore, the approach taken in the static context, in particular that of Wallis (1980, p.63), also breaks down. Finally, depending on the number of shocks in the model, the system can be singular. Singularity implies that the conventional information matrix is not well defined, therefore the rank condition in Rothenberg (1971) can be inapplicable.

Our local identification analysis builds on Qu and Tkachenko (2012). The key idea is that, under indeterminacy, the solutions continue to be representable as a stationary vector linear process whose second order properties continue to be fully characterized by the spectral density matrix $f_\theta(\omega)$ over $\omega \in [-\pi, \pi]$. The elements of $f_\theta(\omega)$ can then be treated as mappings from the finite dimensional augmented parameter space to a space of complex valued functions defined over $[-\pi, \pi]$. Local identification holds if and only if the overall mapping is locally injective. Once this perspective is established, the results reported in this section then follow readily from those obtained in Section 3 in Qu and Tkachenko (2012). Nevertheless, we still present detailed discussions and proofs to ease the understanding and to facilitate the methods' application in practice.

We start with the local identification of both the structural and the sunspot parameters. Intuitively, if this holds, then it is potentially possible to locally determine both the parameters describing technology and preferences (θ^D) and those governing the equilibrium beliefs of agents (θ^U).

Definition 1 *The parameter vector θ is said to be locally identifiable from the second order properties of $\{Y_t\}$ at $\theta = \theta_0$ if there exists an open neighborhood of θ_0 in which $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ necessarily implies $\theta_1 = \theta_0$.*

The above definition can also be stated equivalently in the time domain. Let $\{\Upsilon_\theta(k)\}_{k=-\infty}^{\infty}$ be the autocovariances of $Y_t(\theta)$ implied by the model. Then, θ is said to be locally identifiable from the second order properties of $\{Y_t\}$ at $\theta = \theta_0$ if there exists an open neighborhood of θ_0 in which $\Upsilon_{\theta_1}(k) = \Upsilon_{\theta_0}(k)$ $(k = 0, \pm 1, ...)$ necessarily implies $\theta_1 = \theta_0$.

The mappings associated with the elements of $f_\theta(\omega)$ are infinite dimensional and difficult to analyze directly. However, as observed in Qu and Tkachenko (2012), for local identification it suffices to consider a finite dimensional matrix based on $f_\theta(\omega)$. To state this precisely, we start with the following assumption.

Assumption 3 *Define*

$$G(\theta) = \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial \theta'} \right)^* \left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial \theta'} \right) d\omega \tag{6}$$

and assume θ_0 is a regular point, that is, $G(\theta)$ has a constant rank in an open neighborhood of θ_0.

The next result provides a rank condition for local identification. It generalizes Theorem 1 in Qu and Tkachenko (2012) to allow for indeterminacy.

Theorem 1 (Local identification) *Under Assumptions 1-3, θ is locally identifiable from the second order properties of $\{Y_t\}$ at $\theta = \theta_0$ if and only if $G(\theta_0)$ is nonsingular.*

The dimension of $G(\theta)$ is always $(p+q)$-by-$(p+q)$. It is invariant to the number of equations, observables and shocks in the model. This feature is advantageous for analyzing identification in medium scale DSGE models.

If the regular point assumption in the theorem is dropped, then the condition is still sufficient but no longer necessary for local identification. This assumption can be examined using the method of nonidentification curves proposed in Qu and Tkachenko (2012). First, consider the simple case where $G(\theta_0)$ has one zero eigenvalue. Then, compute a curve that solves the following differential equation in a neighborhood of θ_0

$$\begin{aligned} \frac{\partial \theta(v)}{\partial v} &= c(\theta), \\ \theta(0) &= \theta_0, \end{aligned} \tag{7}$$

where $c(\theta)$ is the orthonormal eigenvector corresponding to the smallest eigenvalue of $G(\theta)$. If θ_0 is a regular point (i.e., $G(\theta)$ has exactly one zero eigenvalue in an open neighborhood of θ_0), then the spectral density $f_{\theta(v)}(\omega)$ along the curve will be identical (see Appendix C, in particular (C.4)) in a neighborhood of θ_0. Thus, if the computed spectral densities start to deviate from $f_{\theta_0}(\omega)$ once θ leaves θ_0, then this implies that θ_0 is not a regular point and that θ is locally identified at θ_0. Next, consider the more complex situation where $G(\theta_0)$ has k zero eigenvalues. Then, proceeding along Steps 1-4 in Qu and Tkachenko (2012, p. 110), we can locate k different parameter subsets, neither of which includes any other as a proper subset, corresponding to these k zero eigenvalues. This further yields k curves each satisfying (7), each with a different $c(\theta)$. Now, recompute the spectral densities using values on the curves. If spectral densities start to deviate from $f_{\theta_0}(\omega)$ along a curve once θ moves away from θ_0, then, again, this implies that θ_0 is not a regular point and the corresponding parameter subset is identified when the rest are held fixed at θ_0. Finally, note that although the regular point assumption is a commonly maintained assumption in the local identification literature, the method of nonidentification curves appears to be the first method that attempts to check its relevance in practice.

Because the nonidentification curve depicts parameter values that are observationally equivalent to θ_0, it can be used to understand the nature of the identification failure. Prior to our work, Arellano, Hansen and Sentana (2012) proposed a method for constructing tests of underidentification by parameterizing the null parameter values as a curve. They also considered curve estimation in a GMM context. In their study, the parameterization of the curve is assumed to be known *a priori*. Here, the parameterization is derived as a result through a differential equation. The works therefore complement each other, and the method here can potentially be a starting point for estimating the curve in finite samples.

We now consider the identification of θ^D without making statements about that of θ^U. Intuitively, if θ^D is locally identifiable, then it is potentially possible to determine the parameters describing technology and preferences, even though those governing the equilibrium beliefs may be unidentifiable.

Definition 2 *The structural parameter vector θ^D is said to be locally partially identifiable from the second order properties of $\{Y_t\}$ at θ_0 if there exists an open neighborhood of θ_0 in which $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ necessarily implies $\theta_1^D = \theta_0^D$.*[1]

Corollary 1 *Under Assumptions 1-3, θ^D is locally partially identifiable from the second order properties of $\{Y_t\}$ at θ_0 if and only if $G(\theta_0)$ and*

$$G^a(\theta_0) = \begin{bmatrix} G(\theta_0) \\ \partial\theta_0^D/\partial\theta' \end{bmatrix}$$

have the same rank.

This result generalizes the first result in Corollary 3 in Qu and Tkachenko (2012) to allow for indeterminacy. It can also be used to check the local identification of a further subset of θ^D without making statements about the rest of θ, simply by replacing θ_0^D with the corresponding parameter subset of interest.

Next, we consider the identification of θ^D conditional on $\theta^U = \theta_0^U$. Intuitively, if θ^D is locally conditionally identifiable, then it is potentially possible to pin down the parameters describing technology and preferences once we select a mechanism for equilibrium belief formation.

[1] Note that, as in Rothenberg (1971, footnote p.586), the definition does not exclude there being two points satisfying $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ and having the θ^D arbitrarily close in the sense of $\left\|\theta_0^D - \theta_1^D\right\| / \|\theta_0 - \theta_1\|$ being arbitrarily small.

Definition 3 *The structural parameter vector θ^D is said to be locally conditionally identifiable from the second order properties of $\{Y_t\}$ at $\theta = \theta_0$ if there exists an open neighborhood of θ_0^D in which $f_{(\theta_1^D,\theta_0^U)}(\omega) = f_{(\theta_0^D,\theta_0^U)}(\omega)$ for all $\omega \in [-\pi, \pi]$ necessarily implies $\theta_1^D = \theta_0^D$.*

Corollary 2 *Under Assumptions 1-3, θ^D is locally conditionally identifiable from the second order properties of $\{Y_t\}$ at θ_0 if and only if*

$$G^D(\theta_0) = \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta^{D\prime}} \right)^* \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta^{D\prime}} \right) d\omega$$

is nonsingular.

This result generalizes the first result in Corollary 4 in Qu and Tkachenko (2012). It can also be used to check the local identification of a further subset of θ^D conditioning on the rest of θ, simply by replacing θ^D with the corresponding parameter subset of interest.

Finally, we consider identification based on a subset of frequencies, e.g., those corresponding to business cycle fluctuations. Let $W(\omega)$ be an indicator function symmetric about zero to select the desired frequencies. The result below focuses on the identification of θ at θ_0, although extensions to partial and conditional local identification can be formulated analogously. Its proof is omitted as it is essentially the same as that of Theorem 1.

Corollary 3 *Let Assumptions 1-3 hold, but with $G(\theta)$ in Assumption 3 replaced by*

$$G^W(\theta) = \int_{-\pi}^{\pi} W(\omega) \left(\frac{\partial \operatorname{vec} f_{\theta}(\omega)}{\partial \theta'} \right)^* \left(\frac{\partial \operatorname{vec} f_{\theta}(\omega)}{\partial \theta'} \right) d\omega.$$

Then, θ is locally identifiable at θ_0 from the second order properties of $\{Y_t\}$ through the frequencies specified by $W(\omega)$ if and only if $G^W(\theta_0)$ is nonsingular.

4 Global identification

This section considers the global identification of θ at θ_0.

Definition 4 *The parameter vector θ is said to be globally identifiable from the second order properties of $\{Y_t\}$ at θ_0 if, within Θ, $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ necessarily implies $\theta_1 = \theta_0$.*

The criterion we propose is based on the Kullback-Leibler distance computed in the frequency domain. To motivate the idea, suppose that there is a realization from the DSGE model, denoted by $Y_t(\theta)$ $(t = 1, ..., T)$. Then, as $T \to \infty$, the Fourier transform at ω satisfies

$$\frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} Y_t(\theta) \exp(-i\omega t) \to^d N_c(0, f_\theta(\omega)),$$

where $N_c(0, f_\theta(\omega))$ denotes a complex valued multivariate normal distribution with mean zero and covariance matrix $f_\theta(\omega)$. Evaluating the limiting distribution at θ_0 and an alternative value θ_1, we obtain $N_c(0, f_{\theta_0}(\omega))$ and $N_c(0, f_{\theta_1}(\omega))$ respectively. The Kullback-Leibler divergence of the second distribution from the first is

$$\frac{1}{2}\{\text{tr}(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - \log\det(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - n_Y\},$$

where n_Y is the dimension of $Y_t(\theta)$. Averaging over $\omega \in [-\pi, \pi]$, we obtain

$$KL(\theta_0, \theta_1) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\{\text{tr}(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - \log\det(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - n_Y\}d\omega. \tag{8}$$

We label this as the Kullback-Leibler distance between two linearized DSGE models with parameters θ_0 and θ_1. More generally, it is the Kullback-Leibler distance between two vector linear processes with spectral densities $f_{\theta_0}(\omega)$ and $f_{\theta_1}(\omega)$.[2]

The Kullback-Leibler distance is a fundamental concept in the statistics and information theory literature. The frequency domain expression (8) was first obtained by Pinsker (1964, p.198, Theorem 10.5.1) as the entropy rate of one vector stationary Gaussian process with respect to another (in our notation, of $Y_t(\theta_0)$ with respect to $Y_t(\theta_1)$). Later, building on his result, Parzen (1983, p. 232, Theorem 2) showed that the autoregressive spectral density estimator satisfies the maximum entropy principle. In relation to the literature, this paper is the first that utilizes the frequency domain expression (8) to convert the Kullback-Leibler distance into a computational device for checking global identification. The feasibility lies in the feature that its expression only depends on the spectral densities of the DSGE model evaluated under θ_0 and θ_1, which are easily computable even under indeterminacy, as discussed in Section 2.

Assumption 4 *The eigenvalues of $f_\theta(\omega)$ are uniformly bounded away from zero for all $\omega \in [-\pi, \pi]$ and all $\theta \in \Theta$.*

The assumption requires the model to be stochastically nonsingular. Otherwise, the analysis can still be applied to nonsingular subsystems (see the discussion following Corollary 4). In some models, such as Smets and Wouters (2007), the vector of observables Y_t contains first differences

[2]The above derivation is kept informal to ease understanding. A more formal derivation for $KL(\theta_0, \theta_1)$ proceeds as follows. Let $L(\theta)$ denote the frequency domain approximate likelihood based on $Y_t^d(\theta)$; for its expression see (14). Then, it can be shown that

$$T^{-1}E(L(\theta_0) - L(\theta_1)) \to KL(\theta_0, \theta_1),$$

where the expectation is taken with respect to θ_0, i.e., assuming $Y_t^d(\theta)$ is generated with $\theta = \theta_0$. $L(\theta)$ can also be replaced by the time domain Gaussian likelihood, although the derivation will be more tedious.

of series that are trend stationary. This makes $f_\theta(\omega)$ singular, but only at the frequency zero. In such situations, the analysis can be carried out using the detrended rather than differenced series as observables. Furthermore, if the analysis is based on a subset of frequencies away from zero (e.g., the business cycle frequencies), then both detrended and differenced series can be considered.

Theorem 2 (Global identification) *Under Assumptions 1, 2 and 4, θ is globally identified from the second order properties of $\{Y_t\}$ at θ_0 if and only if*

$$KL(\theta_0, \theta_1) > 0 \tag{9}$$

for any $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$.

The main function of the theorem is to reduce the problem of checking global identification to minimizing a deterministic function. Specifically, suppose the conditions presented in the previous section show that the parameters are locally identified. Then, to study global identification, we can proceed to check whether

$$\inf_{\theta_1 \in \Theta \setminus B(\theta_0)} KL(\theta_0, \theta_1) > 0, \tag{10}$$

where $B(\theta_0)$ is an open neighborhood of θ_0. Here, $B(\theta_0)$ serves two purposes. First, it excludes from the minimization parameter values that are arbitrarily close to θ_0, at which the KL criterion will be arbitrarily close to zero under continuity. Second, its shape and size can be varied to examine the sensitivity of identification. For example, we can examine how identification improves when successively larger neighborhoods are excluded. This will be illustrated in Section 7.

The criterion $KL(\theta_0, \theta)$ is a deterministic function of θ. Its calculation is simple because $f_\theta(\omega)$ follows from (5) and the integral can be accurately approximated by an average. For solving the minimization problem (10), there is an array of methods available. The incurred computational cost depends on the dimension of θ. If it is high, then the recently developed parallelized global optimization algorithms can be quite useful. More details on implementation, including the case of medium scale models, are provided in Section 7.

Partial and conditional global identification can also be considered. For example, we can consider the global identification of θ^D without making statements about that of θ^U. In this case, the search in Theorem 2 needs to be over the set $\theta_1 \in \Theta$ with $\theta_1^D \neq \theta_0^D$. Or, we can consider the global identification of θ^D conditional on $\theta^U = \theta_0^U$. In this case, we will need to search over the set $\theta_1 \in \Theta$ with $\theta_1^D \neq \theta_0^D$ and $\theta_1^U = \theta_0^U$. More targeted searches can also be implemented depending on the issue of interest. For example, suppose that the focus is on global identification of the

monetary policy rule parameters while fixing the others. Then, one can search over the values of such parameters while keeping the rest fixed. Such investigations can be conducted sequentially to narrow down the source of the identification failure.

The next result considers global identification using a chosen subset of frequencies. Its proof is omitted. In practice, it can be applied jointly with Corollary 3.

Corollary 4 *Define*

$$KL^W(\theta_0,\theta_1)=\frac{1}{4\pi}\int_{-\pi}^{\pi}W(\omega)\{tr(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega))-\log\det(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega))-n_Y\}d\omega. \tag{11}$$

Then, under Assumptions 1, 2 and 4, θ is globally identified at θ_0 based on the frequencies selected by $W(\omega)$ if and only if $KL^W(\theta_0,\theta_1)>0$ for all $\theta_1\in\Theta$ with $\theta_1\neq\theta_0$.

If the DSGE model is singular, Theorem 2 and Corollary 4 can still be applied to its nonsingular subsystems. Let C denote a selection matrix such that $CY_t(\theta)$ is nonsingular for all $\theta\in\Theta$. $KL(\theta_0,\theta_1)$ can be constructed by replacing $f_{\theta_0}(\omega)$ and $f_{\theta_1}(\omega)$ with $Cf_{\theta_0}(\omega)C'$ and $Cf_{\theta_1}(\omega)C'$ respectively. Then, θ is globally identified from the second order properties (or the frequencies selected by $W(\omega)$) of $\{CY_t\}$ at θ_0 if and only if (9) (or (11)) holds. Such an exercise is relevant because, for singular models, likelihood based analysis can only be applied to the nonsingular subsystems. For discussions on selecting observables in singular models one can refer to Guerron-Quintana (2010) and Canova, Ferroni and Matthes (2013).

4.1 Extension: comparing different model structures

In the above analysis, the DSGE models differ only in terms of parameter values θ. In fact, the framework can permit the models to have different structures (e.g., they can have different monetary policy rules, different types of frictions or shock processes, or different determinacy properties; these three exercises will be considered in Section 7). More specifically, suppose $Y_t(\theta)$ and $Z_t(\phi)$ are two vector linear processes generated by two DSGE structures (Structures 1 and 2) with spectral densities $f_\theta(\omega)$ and $h_\phi(\omega)$, where $\theta\in\Theta,\phi\in\Phi$, and Θ and Φ are finite dimensional and compact. Suppose we treat Structure 1 with $\theta=\theta_0$ as the benchmark specification and are interested in whether Structure 2 can generate the same dynamic properties.

Definition 5 *We say Structure 2 is distinct from Structure 1 at $\theta=\theta_0$ if, for any $\phi\in\Phi$, $h_\phi(\omega)\neq f_{\theta_0}(\omega)$ for some $\omega\in[-\pi,\pi]$.*

Corollary 5 *Let Assumptions 1, 2 and 4 hold for $Y_t(\theta)$, $Z_t(\phi)$ and their respective spectral densities. Define*

$$KL_{fh}(\theta,\phi) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\{tr(h_\phi^{-1}(\omega)f_\theta(\omega)) - \log\det(h_\phi^{-1}(\omega)f_\theta(\omega)) - n_Y\}d\omega. \tag{12}$$

Then, Structure 2 is distinct from Structure 1 at $\theta = \theta_0$ if and only if

$$\inf_{\phi\in\Phi} KL_{fh}(\theta_0,\phi) > 0.$$

The proof is the same as Theorem 2 by replacing $f_{\theta_1}(\omega)$ with $h_\phi(\omega)$, therefore omitted. In practice, different model structures may involve overlapping but different sets of observables. To apply the above result, we can specify matrix lag polynomials $C_1(L)$ and $C_2(L)$ such that $C_1(L)Y_t(\theta)$ and $C_2(L)Z_t(\phi)$ return the common observables. Then, $KL_{fh}(\theta,\phi)$ can be constructed by replacing $f_\theta(\omega)$ and $h_\phi(\omega)$ with $C_1(\exp(-i\omega))f_\theta(\omega)C_1(\exp(-i\omega))^*$ and $C_2(\exp(-i\omega))h_\phi(\omega)C_2(\exp(-i\omega))^*$. Also, it is simple to compare model structures based on a chosen subset of frequencies. Define

$$KL_{fh}^W(\theta,\phi) = \frac{1}{4\pi}\int_{-\pi}^{\pi} W(\omega)\{\mathrm{tr}(h_\phi^{-1}(\omega)f_\theta(\omega)) - \log\det(h_\phi^{-1}(\omega)f_\theta(\omega)) - n_Y\}d\omega. \tag{13}$$

Then, Structure 2 is distinct from Structure 1 at $\theta = \theta_0$ within the frequencies selected by $W(\omega)$ if and only if $KL_{fh}^W(\theta_0,\phi) > 0$ for all $\phi \in \Phi$.

The magnitudes of the criterion functions in Theorem 2 and Corollaries 4 and 5 are informative about the strength of identification. Intuitively, if there exists a value of θ distant from θ_0 (or, more generally, a model structure with a very different theoretical underpinning) that makes the criterion functions very close to zero, then in finite samples it can be nearly impossible to distinguish between these two parameter values (or the two structures at some parameter values) on the grounds of their quantitative implications. These points are made precise in the next section.

5 Empirical distance between two models

This section develops a measure for the feasibility of distinguishing a model structure with spectral density $h_{\phi_0}(\omega)$ from a structure with $f_{\theta_0}(\omega)$ constrained by a hypothetical sample size T.

To motivate the construction of the measure, consider the frequency domain approximate Gaussian log likelihoods constructed using a sample of size T. When applied to the two model

structures, they equal (up to a constant addition)

$$\begin{aligned} L_f(\theta_0) &= -\frac{1}{2}\sum_{j=1}^{T-1}\left\{\log\det(f_{\theta_0}(\omega_j)) + \mathrm{tr}(f_{\theta_0}^{-1}(\omega_j)I(\omega_j))\right\}, \\ L_h(\phi_0) &= -\frac{1}{2}\sum_{j=1}^{T-1}\left\{\log\det(h_{\phi_0}(\omega_j)) + \mathrm{tr}(h_{\phi_0}^{-1}(\omega_j)I(\omega_j))\right\}, \end{aligned} \tag{14}$$

where $\omega_j = 2\pi j/T$ $(j = 1, 2, ..., T-1)$ are the Fourier frequencies and $I(\omega_j)$ denotes the periodogram

$$\begin{aligned} I(\omega_j) &= w(\omega_j)\, w(\omega_j)^*, \\ w(\omega_j) &= \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T} Y_t \exp(-i\omega_j t). \end{aligned}$$

The log likelihood ratio for testing $f_{\theta_0}(\omega)$ against $h_{\phi_0}(\omega)$ after division by T can be decomposed as

$$\begin{aligned} \frac{1}{T}(L_h(\phi_0) - L_f(\theta_0)) &= \frac{1}{2T}\sum_{j=1}^{T-1}\left\{\log\det(h_{\phi_0}^{-1}(\omega_j)f_{\theta_0}(\omega_j)) - \mathrm{tr}(h_{\phi_0}^{-1}(\omega_j)f_{\theta_0}(\omega_j)) + n_Y\right\} \\ &\quad + \frac{1}{2T}\sum_{j=1}^{T-1}\mathrm{tr}\left\{\left(f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j)\right)(I(\omega_j) - f_{\theta_0}(\omega_j)\right\}. \end{aligned}$$

The first summation on the right hand side converges to $-KL_{fh}(\theta_0, \phi_0)$. The second term, after multiplication by $T^{1/2}$, satisfies a central limit theorem under the null hypothesis that the sample has the spectral density $f_{\theta_0}(\omega)$. A similar decomposition can be used to analyze the log likelihood ratio under the alternative hypothesis. In order to formally state these distributional results, we impose the following assumption.

Assumption 5 *The elements of $f_\theta(\omega)$ and $h_\phi(\omega)$ belong to the Lipschitz class of degree $\beta > 1/2$ with respect to ω.*[3] *Let $\widetilde{\varepsilon}_{ta}$ denote the a-th element of $\widetilde{\varepsilon}_t = (\varepsilon_t', \epsilon_t')$. Assume the joint fourth cumulant satisfies* $\mathrm{cum}(\widetilde{\varepsilon}_{ta}, \widetilde{\varepsilon}_{sb}, \widetilde{\varepsilon}_{uc}, \widetilde{\varepsilon}_{vd}) = 0$ *for all $1 \le a, b, c, d \le \dim(\widetilde{\varepsilon}_t)$ and $-\infty < t, s, u, v < \infty$.*

The first part of the assumption requires sufficient smoothness in the spectral densities with respect to ω. This is a very mild assumption in the context of DSGE models. For illustration, consider $f_\theta(\omega)$ and write the lag polynomial $H(L;\theta)$ in (3) explicitly as $\sum_{j=0}^{\infty}\rho_j(\theta)L^j$. Then, the stated Lipschitz condition is satisfied if $\sum_{j=0}^{\infty} j^\beta ||\rho_j(\theta)|| \le \infty$ (Hannan, 1970, p. 311-312), which typically holds because (4) implies that $\rho_j(\theta)$ decays exponentially. The joint cumulant condition is

[3] Let $g(\omega)$ be a scalar valued function defined on an interval B. We say that g belongs to the Lipschitz class of degree β if there exists a finite constant M such that $|g(\omega_1) - g(\omega_2)| \le M|\omega_1 - \omega_2|^\beta$ for all ω_1, $\omega_2 \in B$.

more stringent. It is satisfied if the errors are distributed normally. The latter covers the majority of models studied in the DSGE literature. The relaxation of this assumption will be considered later in the paper.

Theorem 3 *Let Assumptions 1, 2, 4 and 5 hold for both $f_\theta(\omega)$ and $h_\phi(\omega)$. Then, under the null hypothesis that $f_{\theta_0}(\omega)$ is the true spectral density,*

$$T^{1/2}\left(\frac{1}{T}\left(L_h(\phi_0)-L_f(\theta_0)\right)+KL_{fh}(\theta_0,\phi_0)\right)\to^d N(0,V_{fh}(\theta_0,\phi_0)).$$

Under the alternative hypothesis that $h_{\phi_0}(\omega)$ is the true spectral density,

$$T^{1/2}\left(\frac{1}{T}\left(L_h(\phi_0)-L_f(\theta_0)\right)-KL_{hf}(\phi_0,\theta_0)\right)\to^d N(0,V_{hf}(\phi_0,\theta_0)).$$

where $KL_{fh}(\theta_0,\phi_0)$ is given in (12), $KL_{hf}(\phi_0,\theta_0)$ is as in (12) but with h_ϕ and f_θ reversed,

$$\begin{aligned}V_{fh}(\theta_0,\phi_0) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{\left[I-f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)\right]\left[I-f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)\right]\right\}d\omega,\\ V_{hf}(\phi_0,\theta_0) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{\left[I-h_{\phi_0}(\omega)f_{\theta_0}^{-1}(\omega)\right]\left[I-h_{\phi_0}(\omega)f_{\theta_0}^{-1}(\omega)\right]\right\}d\omega.\end{aligned}$$

Theorem 3 can be viewed as a frequency domain counterpart to Cox's (1961) results on non-nested hypothesis testing. Different from the usual situation, here both the null and alternative distributions of $T^{-1/2}\left(L_h(\phi_0)-L_f(\theta_0)\right)$ are readily computable from the spectral densities of the models. Consequently, the theorem permits straightforward calculation of the approximate power of the standardized likelihood ratio test $T^{-1/2}\left(L_h(\phi_0)-L_f(\theta_0)\right)$ of $f_{\theta_0}(\omega)$ against $h_{\phi_0}(\omega)$. To proceed, select a significance level α. The first result (the null limiting distribution) in the theorem implies that the critical value for the test equals, up to $o(1)$,

$$q_\alpha=-T^{1/2}KL_{fh}(\theta_0,\phi_0)+\sqrt{V_{fh}(\theta_0,\phi_0)}z_{1-\alpha},$$

where $z_{1-\alpha}$ is the $100(1-\alpha)th$ percentile of the standard normal distribution. The second result (the limiting distribution under the alternative) implies that the testing power, up to $o(1)$, equals

$$p_{fh}(\theta_0,\phi_0,\alpha,T)=\Pr\left(Z>\frac{q_\alpha-T^{1/2}KL_{hf}(\phi_0,\theta_0)}{\sqrt{V_{hf}(\phi_0,\theta_0)}}\right),$$

where $Z\sim N(0,1)$.

We term $p_{fh}(\theta_0,\phi_0,\alpha,T)$ as the measure of the empirical distance of the model with spectral density $h_{\phi_0}(\omega)$ from the one with $f_{\theta_0}(\omega)$. To obtain it, the main work is in computing $KL_{fh}(\theta_0,\phi_0)$,

$KL_{hf}(\phi_0, \theta_0)$, $V_{fh}(\theta_0, \phi_0)$ and $V_{hf}(\phi_0, \theta_0)$. They depend only on the spectral densities $f_{\theta_0}(\omega)$ and $h_{\phi_0}(\omega)$ without any reference to any data. Computing them thus only requires solving the two models once to compute the respective spectral densities. No simulation is needed. Also, there is no need to write down any likelihood. The other quantities in $p_{fh}(\theta_0, \phi_0, \alpha, T)$ refer only to the $N(0,1)$ distribution and are trivial to obtain.

The measure $p_{fh}(\theta_0, \phi_0, \alpha, T)$ has the following features. (1) For any T and α, its value is always between 0 and 1. It is increasing in $KL_{fh}(\theta_0, \phi_0)$ and $KL_{hf}(\phi_0, \theta_0)$, thus a higher value signifies that it is easier to distinguish between the two model structures. (2) If $KL_{fh}(\theta_0, \phi_0) > 0$ (so will be $KL_{hf}(\phi_0, \theta_0)$), then the measure will be strictly increasing in T and approach 1 as $T \to \infty$.

As with the KL criterion, the empirical distance measure can be applied to structures with overlapping but different sets of observables. This in particular permits measuring the distance between a small and a medium scale DSGE model. It can also be computed using a subset of frequencies. In this case, $KL_{fh}(\theta_0, \phi_0)$ and $KL_{hf}(\phi_0, \theta_0)$ need to be replaced with $KL_{fh}^W(\theta_0, \phi_0)$ and $KL_{hf}^W(\phi_0, \theta_0)$ (see (13) for their definition) and $V_{fh}(\theta_0, \phi_0)$ and $V_{hf}(\phi_0, \theta_0)$ with

$$\begin{aligned} V_{fh}^W(\theta_0, \phi_0) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\omega) \operatorname{tr} \left\{ \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \right\} d\omega, \\ V_{hf}^W(\phi_0, \theta_0) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\omega) \operatorname{tr} \left\{ \left[I - h_{\phi_0}(\omega) f_{\theta_0}^{-1}(\omega) \right] \left[I - h_{\phi_0}(\omega) f_{\theta_0}^{-1}(\omega) \right] \right\} d\omega. \end{aligned}$$

We label the resulting measure as

$$p_{fh}^W(\theta_0, \phi_0, \alpha, T).$$

This measure is valuable as it can decompose to what extent the quantitative differences between models are driven by, for example, the business cycle frequencies as opposed to others.

When the requirement $\operatorname{cum}(\widetilde{\varepsilon}_{ta}, \widetilde{\varepsilon}_{sb}, \widetilde{\varepsilon}_{uc}, \widetilde{\varepsilon}_{vd}) = 0$ in Assumption 6 is relaxed, the asymptotic variances in the theorem will depend on the joint fourth cumulant of the structural shocks. We state this as a corollary.

Corollary 6 *Let Assumptions 1, 2, 4 and 5 hold for both $f_\theta(\omega)$ and $h_\phi(\omega)$, but with the cumulant condition in Assumption 5 replaced by*

$$\operatorname{cum}(\widetilde{\varepsilon}_{ta}, \widetilde{\varepsilon}_{sb}, \widetilde{\varepsilon}_{uc}, \widetilde{\varepsilon}_{vd}) = \begin{cases} \kappa_{abcd} & \text{if } t = s = u = v, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the results in Theorem 3 still hold after redefining $V_{fh}(\theta_0, \phi_0)$ *and* $V_{hf}(\phi_0, \theta_0)$ *as*

$$\begin{aligned} V_{fh}(\theta_0, \phi_0) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \right\} d\omega + M_{fh}(\theta_0, \phi_0), \\ V_{hf}(\phi_0, \theta_0) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ \left[I - h_{\phi_0}(\omega) f_{\theta_0}^{-1}(\omega) \right] \left[I - h_{\phi_0}(\omega) f_{\theta_0}^{-1}(\omega) \right] \right\} d\omega + M_{hf}(\theta_0, \phi_0), \end{aligned}$$

where the (j,l)-th element of $M_{fh}(\theta_0, \phi_0)$ *equals*

$$\begin{aligned} &\left(\frac{1}{4\pi} \right)^2 \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} H^*(\omega) \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) H(\omega) d\omega \right]_{ab} \\ &\times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} H^*(\omega) \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) H(\omega) d\omega \right]_{cd}. \end{aligned}$$

In the above $[.]_{ab}$ *denotes the (a,b)-th element of the matrix,* κ_{abcd} *denotes the joint cumulant in the null model,* $H(\omega)$ *equals* $H(\exp(-i\omega); \theta_0)$ *in (5), and* $H^*(\omega)$ *is its conjugate transpose. The (j,l)-th element of* $M_{hf}(\phi_0, \theta_0)$ *satisfies the same expression, but with* κ_{abcd} *now denoting the joint cumulant in the alternative model and* $H(\omega)$ *determined by the polynomial operator in* $h_{\phi_0}(\omega)$.

The result implies that, when computing the empirical distance measure under non-Gaussian innovations, only the asymptotic variances need to be modified. For example, if the model features Student-t shocks, then the main additional work is to compute κ_{abcd} in terms of the degrees of freedom parameters. The earlier comments about the overlapping but different sets of observables and subsets of frequencies are also still valid. For the former, $H(\omega)$ and the spectral densities in $M_{fh}(\theta_0, \phi_0)$ and $M_{hf}(\phi_0, \theta_0)$ will need to be multiplied by the selection matrices, while for the latter the integrals also need to be adjusted accordingly.

A range of issues can be tackled using the empirical distance measure. First, the value of ϕ can be varied to measure the empirical distance between a family of alternative models and the benchmark model, e.g., computing for some pre-specified set S

$$\inf_{\phi \in S} p_{fh}(\theta_0, \phi, \alpha, T) \quad \text{or} \quad \sup_{\phi \in S} p_{fh}(\theta_0, \phi, \alpha, T).$$

Second, the measure can be treated as a function of T to illustrate to what extent distinguishing models becomes feasible as T increases. Third, a family of measures can be computed by allowing for different departures from the benchmark model, e.g., by shutting down different types of real or nominal frictions. This in turn will be useful for understanding which modeling features are essential for generating the dynamics of the model. Some examples of such applications are reported in Section 7.

6 Application to (factor augmented) VARMA models

This section highlights the application of the proposed methods to (factor augmented) VARMA models. Consider

$$\begin{aligned} A(L)Y_t &= \lambda(L)f_t + B(L)\varepsilon_t, \\ f_t &= \Gamma(L)f_{t-1} + \zeta_t, \end{aligned}$$

where Y_t is an n_Y-by-1 vector of observables, f_t contains the latent factors, ε_t and ζ_t are two sequences of serially uncorrelated structural shocks satisfying $Var(\varepsilon_t) = \Sigma, Var(\zeta_t) = I$ and $E\varepsilon_t\zeta_s' = 0$ for all t and s. $A(L), B(L)$, $\lambda(L)$ and $\Gamma(L)$ are finite order matrix lag polynomials. If $\lambda(L) = 0$, then the model reduces to a VARMA model. If $B(L) = I$, it becomes a factor augmented VAR. If $\lambda(L) = 0$ and $B(L) = I$, then it is simply a structural VAR.

Assume all the roots of $|A(z)| = 0$ lie outside of the unit circle. Let θ denote the vector of the structural parameters in the model. Then, Y_t has the following vector moving average representation

$$Y_t = H_1(L;\theta)\varepsilon_t + H_2(L;\theta)\zeta_t, \tag{15}$$

where

$$H_1(L;\theta) = A(L)^{-1}B(L), \;\; H_2(L;\theta) = A(L)^{-1}\lambda(L)[I - \Gamma(L)L]^{-1}.$$

Its spectral density equals

$$f_\theta(\omega) = \frac{1}{2\pi}H_1(\exp(-i\omega);\theta)\Sigma H_1(\exp(-i\omega);\theta)^* + \frac{1}{2\pi}H_2(\exp(-i\omega);\theta)H_2(\exp(-i\omega);\theta)^*.$$

Some of the identification restrictions, such as the diagonality of Σ or factor loading restrictions on elements of $\lambda(L)$ can be directly incorporated, while other forms of restrictions, such as the long run restrictions, can be written as constraints on θ. Let $\psi(\theta) = 0$ be the collection of such constraints one wishes to impose. Theorem 2 can then be used for checking global identification. There, global identification at θ_0 holds if and only if

$$KL(\theta_0, \theta_1) > 0$$

for all $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$ satisfying $\psi(\theta_1) = 0$. The procedures in Section 5 are also applicable. For example, we can contrast models with and without factor augmentations, models with different identification restrictions, or comparing a (factor augmented) VARMA with a DSGE model.

7 Illustrative applications

This section applies the developed methods to three models. The first two are small scale models studied in An and Schorfheide (AS, 2007) and Lubik and Schorfheide (LS, 2004). The third is the medium scale DSGE model of Smets and Wouters (SW, 2007). Putting these three models together allows us to compare how identification properties can subtly differ between similar small scale models and change when moving to a medium scale model. Throughout the analysis, we treat determinacy and indeterminacy within a given model as two different structures. This ensures the differentiability of the spectrum as required in Assumption 2.

The main algorithms for optimization are tailored versions of the genetic algorithm and constrained local minimization algorithm of Matlab. They are implemented on an eight-core Dell Precision desktop. For the small scale models, it typically takes less than one hour to obtain, say, a column reported in Table 2. For the medium scale model it takes a few hours. All results reported are verified by running the algorithms multiple times. The replication files are available upon request and will be included as an online supplement.

7.1 An and Schorfheide (2007)

Previously, Qu and Tkachenko (2012) considered this model's local identification properties under determinacy. In particular, they showed that the parameters in the Taylor rule are locally unidentified at a particular value and obtained parameter values that yielded observational equivalence. This subsection conducts a further analysis under indeterminacy. First, it examines local and global identification under indeterminacy. Next, it compares behavioral implications of alternative monetary policy rules. Then, it illustrates how the proposed empirical distance measure can be used to gauge the feasibility of distinguishing between different model specifications. Finally, it reexamines some results when only business cycle frequencies are used for identification.

The log linearized model is:

$$\begin{aligned} y_t &= E_t y_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}(r_t - E_t\pi_{t+1} - E_t z_{t+1}), \\ \pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - g_t), \\ r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_1\pi_t + (1-\rho_r)\psi_2(y_t - g_t) + \varepsilon_{rt}, \\ g_t &= \rho_g g_{t-1} + \varepsilon_{gt}, \\ z_t &= \rho_z z_{t-1} + \varepsilon_{zt}, \end{aligned} \tag{16}$$

where y_t denotes output, π_t is inflation, r_t is the nominal interest rate, g_t is government spending,

z_t is a disturbance in the technology growth and $\varepsilon_{rt} \sim N(0, \sigma_r^2)$, $\varepsilon_{gt} \sim N(0, \sigma_g^2)$ and $\varepsilon_{zt} \sim N(0, \sigma_z^2)$ are serially and mutually uncorrelated structural shocks. The structural parameter vector is

$$\theta^D = (\tau, \beta, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z)'.$$

The vector of state variables is $S_t = (\pi_t, y_t, r_t, g_t, z_t, E_t\pi_{t+1}, E_t y_{t+1})'$ and the vector of observables is $Y_t = (r_t, y_t, \pi_t)'$. In this model, and also in Lubik and Schorfheide (2004), the equilibrium is determinant if and only if $\psi_1 > 1 - (1-\beta)\psi_2/\kappa$. Because the value of β is very close to one, $\psi_1 = 1$ constitutes an approximate boundary. Under indeterminacy, the sunspot shock (ϵ_t) is one dimensional and is allowed to be correlated with the fundamental shocks: $\epsilon_t = M_{r\epsilon}\varepsilon_{rt} + M_{g\epsilon}\varepsilon_{gt} + M_{z\epsilon}\varepsilon_{zt} + \widetilde{\epsilon}_t$, where $\widetilde{\epsilon}_t$ is uncorrelated with the fundamental shocks with standard deviation σ_ϵ. Consequently, the sunspot parameter vector is

$$\theta^U = (M_{r\epsilon}, M_{g\epsilon}, M_{z\epsilon}, \sigma_\epsilon)'. \tag{17}$$

Let

$$\theta = \begin{pmatrix} \theta^D \\ \theta^U \end{pmatrix}.$$

7.1.1 Local identification

In order to obtain empirically relevant values for θ under indeterminacy, the model is estimated with the dataset from Smets and Wouters (2007) for the period 1960:I-1979:II (In the literature, this period is often regarded as a regime with indeterminacy arising from a passive monetary policy.) using the frequency domain quasi-Bayesian procedure based on the second order properties suggested in Qu and Tkachenko (2012). The prior distributions are taken from Table 1 in Lubik and Schorfheide (2004). When minimizing the log-posterior, β is reparameterized as $r^* = 400(1/\beta - 1)$ with r^* representing the annualized steady state real interest rate. The results are reported in Table 1. Later on, when computing the rank of $G(\theta_0)$, we use the Matlab default tolerance level: $tol = size(G)eps(\|G\|)$, where eps is the floating point precision of G. All the results are then verified with alternative methods.

First, consider local identification at the posterior mean:

$$\theta_0 = (\underbrace{2.51, 0.995, 0.49, 0.63, 0.23, 0.87, 0.66, 0.60, 0.27, 0.58, 0.62}_{\theta^D}, \underbrace{0.53, -0.06, 0.26, 0.19}_{\theta^U})'.$$

The smallest eigenvalue of $G(\theta_0)$ obtained by applying Theorem 1 equals 4.5E-05, while the tolerance level is at a much smaller value of 6.8E-12. This suggests that θ_0 is locally identified. This

result is further verified in two ways. First, we compute the curve defined in (7) with $c(\theta)$ determined by the smallest eigenvalue. The step size is set to 1E-04. The following three measures, suggested in Qu and Tkachenko (2012), are used to gauge the differences between $f_\theta(\omega)$ and $f_{\theta_0}(\omega)$ along the curve (Let $f_{\theta hl}(\omega)$ denote the (h,l)-th element of the spectral density matrix. Note that when computing the second measure, the denominator is evaluated at the same frequency that maximizes the numerator.)

$$\begin{aligned} \text{Maximum absolute deviation:} \quad & \max_{\omega\in[0,\pi]} \mid f_{\theta hl}(\omega) - f_{\theta_0 hl}(\omega)\mid, \\ \text{Maximum absolute deviation in relative form} \quad : \quad & \frac{\max_{\omega\in[0,\pi]} \mid f_{\theta hl}(\omega) - f_{\theta_0 hl}(\omega)\mid}{\mid f_{\theta_0 hl}(\omega)\mid}, \\ \text{Maximum relative deviation} \quad : \quad & \max_{\omega\in[0,\pi]} \frac{\mid f_{\theta hl}(\omega) - f_{\theta_0 hl}(\omega)\mid}{\mid f_{\theta_0 hl}(\omega)\mid}. \end{aligned} \tag{18}$$

The three measures equal 5.5E-04, 2.4E-04 and 1.0E-2 respectively when $||\theta - \theta_0||$ reaches 0.045; they further increase and all exceed 1.0E-2 when $||\theta-\theta_0||$ reaches 1.30. These values are significantly higher than what would arise from purely numerical errors associated with the Euler approximation method for the curve: the cumulative error should be of the same order as the step size used, which here equals 1.0E-04. Second, we compute the empirical distance measure $p_{ff}(\theta_0, \theta, 0.05, T)$ using the above θ satisfying $||\theta - \theta_0|| = 0.045$. The values are $0.0524, 0.0531, 0.0535$ and 0.0576 for T=80,150,200 and 1000 respectively. They are all above 5% and are increasing with sample size. Both methods confirm that the model is locally identified at θ_0.

Next, we draw parameter values from the posterior distribution in Table 1 and check the local identification at each point. The following bounds on parameter values are imposed: $\tau \in [0.01, 10]$, $\beta \in [0.9, 0.999]$, $\kappa \in [0.01, 5]$, $\psi_1 \in [0.01, 0.9]$, $\psi_2 \in [0.01, 5]$, $\rho_r \in [0.1, 0.99]$, $\rho_g \in [0.1, 0.99]$, $\rho_z \in [0.1, 0.99]$, $\sigma_r \in [0.01, 3]$, $\sigma_g \in [0.01, 3]$, $\sigma_z \in [0.01, 3]$, $M_{r\epsilon} \in [-3, 3]$, $M_{g\epsilon} \in [-3, 3]$, $M_{z\epsilon} \in [-3, 3]$, $\sigma_\epsilon \in [0.01, 3]$. Out of the 4000 draws, the smallest eigenvalues are consistently above the tolerance levels (i.e., implying the parameters are locally identified), except for two cases. The two cases involve ρ_r equal to 0.983 and 0.987, which are close to the boundary value of 0.99. Further, for these two points, the absolute and relative deviation measures increase noticeably along the curves, all exceeding E-02 when $\|\theta - \theta_0\|$ reaches 0.050 and 0.709 respectively. For these values the empirical distance measures for T=1000 equal 0.0524 and 0.0625. This suggests that the model is in fact identified at these two parameter values. Therefore, the local identification property is not confined to a particular parameter value, but rather is a generic feature of this model under indeterminacy.

The above identification feature is in sharp contrast with that under determinacy. Previously, Qu and Tkachenko (2012) considered a parameter value taken from Table 3 in An and Schorfheide (2007): $\theta^D = (2,\ 0.9975,\ 0.33,\ 1.5,\ 0.125,\ 0.75,\ 0.95,\ 0.9,\ 0.2,\ 0.6,\ 0.3)'$. Below, we further consider alternative parameter values. The posterior mean obtained using the post-1982 subsample equals (reported in the second last column in Table 1)

$$\theta_0^D = (2.24, 0.995, 0.84, 2.32, 0.26, 0.65, 0.93, 0.88, 0.23, 0.77, 0.26)'. \tag{19}$$

At this value, the smallest eigenvalue of $G(\theta_0^D)$ equals 6.5E-13, below the corresponding default tolerance level 4.0E-11. This suggests that θ_0^D is locally unidentified. This is further confirmed by the deviation and empirical distance measures. The deviation measures remain below E-05 after $\left\|\theta^D - \theta_0^D\right\|$ reaches 1. The empirical distance measure equals 0.0500 at this θ^D when T=1000. Furthermore, because all the other eigenvalues are substantially above the tolerance level (the second smallest eigenvalue equals 5.7E-05), only one subset of parameters is responsible for the identification failure. They correspond to the parameters in the Taylor rule: $(\psi_1, \psi_2, \rho_r, \sigma_r)$. This is consistent with the finding in Qu and Tkachenko (2012). Next, as in the indeterminacy case, we draw parameter values from the posterior distribution in Table 1 and apply Theorem 1 to each point. The same parameter bounds are used, except $\psi_1 \in [1.1, 5]$ and the elements in θ^U are no longer present. Out of the 4000 draws, there are 3998 cases with the smallest eigenvalue below the default tolerance level (i.e., signaling identification failure). Even in the remaining two cases, the eigenvalues are very small, both being of order E-10, and barely exceed the tolerance levels. Also, in both cases, the values of the two measures along the curves (7) remain negligible, the largest still being of order E-05 after $\left\|\theta^D - \theta_0^D\right\|$ reaches 1. This shows that the two parameter values are also locally unidentified. In addition, in all cases considered, the lack of identification is caused by the four parameters in the Taylor rule.

Such a feature that some parameters are identified under indeterminacy but not under determinacy has been documented in the literature (see, e.g., Beyer and Farmer, 2004, Lubik and Schorfheide, 2004). In particular, Lubik and Schorfheide (2004) illustrates analytically that the sunspot fluctuations generate additional dynamics and therefore contribute to parameter identification. Although the literature has only reported this feature in simple models that can be solved analytically, this paper shows that this can also occur in models of empirical relevance.

7.1.2 Global identification under indeterminacy

This subsection considers global identification properties of θ at θ_0. We search for models closest to θ_0 according to the KL criterion over the parameter space excluding a small neighborhood of θ_0, that is, over

$$\{\theta : |\theta - \theta_0|_\infty \geq c\}, \tag{20}$$

where $|.|_\infty$ returns the maximum absolute difference between the elements of θ and θ_0 and c can be set to different values to exclude neighborhoods of different sizes. We consider $c = 0.1, 0.5$ and 1.0. In practice, other definitions of neighborhoods and c can also be considered. For example, we can use the Euclidean distance $\{\theta : \|\theta - \theta_0\| \geq c\}$ or percentage difference $\{\theta : |(\theta - \theta_0)./\theta_0|_\infty \geq c\}$. Then, only the program for optimization needs to be modified accordingly.

The parameter vectors minimizing the KL criterion for $c = 0.1$, 0.5 and 1.0 are reported in Panel (a) in Table 2, where the bold values signify the parameters that change the most. The corresponding KL values and empirical distance measures are reported in Table 3.

First, consider the case with $c = 0.1$. As shown in the third column in Table 2, τ increases by 0.1 to make the constraint binding. The KL criterion, shown in the second column in Table 3, equals 5.36E-07. As this is a small value, we compute the empirical distances $p_{ff}(\theta_0, \theta_{0.1}, 0.05, T)$ for T=80,150,200,1000 to gain more insight into the nature of identification. The values are all above 0.05 and increase consistently with the sample size. This confirms that the two models are not observationally equivalent. Meanwhile, it also shows that it will be very hard to distinguish between the two models with typical sample sizes, as the distance is only 0.0509 for T=80 and increases just to 0.0534 when T=1000. Overall, the findings imply: (1) the model is globally identified at θ_0, (2) the model with $\theta = \theta_{0.1}$ is difficult to distinguish from $\theta = \theta_0$ in practice, and (3) the model minimizing the KL criterion is obtained by mainly shifting τ, and small adjustments in ψ_1, ψ_2, σ_g and σ_z, while the only sunspot parameter that noticeably changes is $M_{r\epsilon}$.

Next, consider the case with $c = 0.5$. The KL criterion equals 1.57E-05. The empirical distance reported in Table 3 equals 0.0553 at T=80 and 0.0710 at T=1000. This suggests that it is still hard to distinguish these two models. As in the case with $c = 0.1$, the most notable change is in τ, which increases by 0.5 compared to its value at θ_0. The parameters ψ_1, σ_z, $M_{r\epsilon}$ continue increasing, while ψ_2 further decreases. Additionally, more sunspot parameters change, as $M_{g\epsilon}$ goes down and $M_{z\epsilon}$ increases slightly.

Now, consider $c = 1.0$. The empirical distance measure equals 0.0621 when T=80 and gradually climbs to 0.1041 at T=1000. Therefore, it is still hard to differentiate between the two models in

empirical work. As before, the largest change in the parameter vector is the unit increase in τ from its initial θ_0 value. Most of the remaining parameters also change, albeit in much smaller magnitudes, to maintain the closest KL criterion to the model at θ_0.

In summary, no observationally equivalent model is found when some neighborhood around θ_0 is excluded from the search. This implies that the model is globally identified at θ_0. Still, the models generated by $\theta_{0.1}$, $\theta_{0.5}$ and $\theta_{1.0}$ produce dynamics similar to those at θ_0 and will be difficult to differentiate from the latter in typical sample sizes.

The analysis shows that an important channel for near observational equivalence is the weak identification of τ. We now assess whether the identification improves substantially if τ is kept fixed at its original value of 2.51. The parameter values minimizing the KL criterion are reported in Panel (b) in Table 2. At these values, the KL criteria (see Table 3) equal 1.04E-05, 1.65E-04 and 3.35E-04 respectively. The empirical distance measures are consistent with the relatively mild increase in the KL criteria. In particular, when $c = 1.0$, the empirical distances are 0.0769 and 0.0900 when T=80 and 150. Thus, distinguishing between models remains a challenge. Interestingly, in all three cases, the parameter that moves the most is the output gap target parameter ψ_2. This implies that different policy rule parameters can result in near observational equivalence even if they are globally identified.

Finally, we fix both τ and ψ_2 at their original values. The improvement in identification strength is now much more pronounced. For $c = 0.5$, KL equals 4.24E-04 and empirical distances are 0.0839 and 0.2339 for T=80 and 1000. For $c = 1.0$, KL reaches 1.20E-3 and the empirical distances are 0.1147 and 0.4648 respectively. Therefore, after fixing both parameters, distinguishing between alternative models is now more feasible. This process can be continued by fixing additional parameters, to further measure the improvements in the strength of identification.

The procedure discussed in this subsection not only allows us to check global identification and measure the feasibility of distinguishing between models, it is also useful for sequentially extracting parameters responsible for (near) identification failure. This will be further illustrated when considering the next two models.

7.1.3 Identification of policy rules

This subsection considers the feasibility of distinguishing the model (16) from two models that have the same structure as (16) except for the monetary policy rule. In the first case, the central bank

responds to expected inflation rather than current inflation:

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1 E_t\pi_{t+1} + (1-\rho_r)\psi_2(y_t - g_t) + \varepsilon_{rt}. \tag{21}$$

In the second, the central bank reacts to output growth instead of output gap:

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1 \pi_t + (1-\rho_r)\psi_2(\Delta y_t + z_t) + \varepsilon_{rt}.$$

The latter rule is also considered in An and Schorfheide (2007). In each case, denote the model's parameter vector by ϕ (elements ordered in the same way as in θ) and its spectral density by $h_\phi(\omega)$. Throughout the analysis, the parameter values for the original model are fixed at the posterior means, while those for the alternative models are determined through minimizing the KL criterion. The results are summarized in Table 4.

First, consider the expected inflation rule. Under indeterminacy, $KL_{fh}(\theta_0, \phi)$ is minimized at

$$\phi = (2.51, 0.995, 0.49, 0.63, 0.54, 0.87, 0.66, 0.60, 0.27, 0.58, 0.62, 0.53, -0.06, 0.26, 0.19)', \tag{22}$$

where only the value of ψ_2 changes relative to θ_0. The resulting KL criterion equals 3.38E-14. Its small magnitude suggests that the two models are observationally equivalent. To further verify this, we compute the empirical distance measures for the four sample sizes. They stay at 0.0500 even after the sample size is increased to 1000. This confirms that the expected inflation rule under (22) indeed yields dynamics identical to those of the baseline inflation rule under θ_0. We now consider whether such equivalence also arises under determinacy at $\theta^D = \theta_0^D$. The KL criterion is minimized at

$$\phi^D = (2.24, 0.995, 0.84, 2.30, 2.25, 0.65, 0.93, 0.88, 0.23, 0.77, 0.26)',$$

where the only values different from θ_0^D are ψ_1 and ψ_2. The KL equals 4.24E-14. Again, the empirical distance measure stays at 0.0500 when the sample size is increased to 1000. Thus, policy equivalence also arises under determinacy. This equivalence is intuitive because the Phillips curve implies a deterministic relationship between $\pi_t, E_t\pi_{t+1}$ and $(y_t - g_t)$. Nevertheless, documenting it plays a useful part in illustrating the proposed method at work.

Next, consider the output growth rule. Under indeterminacy, $KL_{fh}(\theta_0, \phi)$ is minimized at

$$\phi = (2.80, 0.997, 0.48, 0.69, 0.01, 0.86, 0.66, 0.59, 0.27, 0.57, 0.63, 0.55, -0.07, 0.27, 0.19)'. \tag{23}$$

The KL criterion equals 1.94E-05. The empirical distance measure equals 0.0560 when the sample size is 80, and slowly increases to 0.0738 when the sample size is 1000. These results suggest that

the two models are not observationally equivalent. However, it is very hard to distinguish between them in typical sample sizes. Under determinacy, $KL_{fh}(\theta_0^D, \phi^D)$ is minimized at

$$\phi^D = (2.17, 0.999, 0.82, 2.37, 0.01, 0.64, 0.93, 0.88, 0.22, 0.77, 0.26)'. \tag{24}$$

The resulting KL criterion equals 6.00E-05. The empirical distance measure equals 0.0614 and 0.0976 for sample sizes of 80 and 1000. Overall, the results are similar to the indeterminacy case. Therefore, in both cases, the output growth rule leads to a model that is nearly but not exactly observationally equivalent to the original model.

7.1.4 Further illustrations of the empirical distance measure

This subsection further illustrates the informativeness of the empirical distance measure by considering a range of models different from (16). Throughout the analysis, (16) is regarded as the default specification and the significance level for the empirical distance measure is set to 0.05.

First, consider a situation where the alternative model is known to be difficult to distinguish from the default model even with a large sample size. Specifically, we take a parameter vector from the nonidentification curve computed previously for θ_0^D in Subsection 7.1.1, such that $\left\|\theta^D - \theta_0^D\right\| = 1$, but truncate the values of the non-identified parameters $(\psi_1, \psi_2, \rho_r, \sigma_r)$ to leave 2 decimal places: $(2.06, 1.23, 0.68, 0.24)$. As expected, the KL criterion is small (1.04E-04), and the value of the empirical distances equal 0.0662, 0.0725, 0.0763 and 0.1193 for T=80, 150, 200 and 1000. This exercise is interesting as it illustrates the magnitude of the empirical distance measure that one could expect when the models are nearly observationally equivalent.

Second, consider the case where all of the parameter values in θ^D are the same as in θ_0^D except for the discount factor β, which is now lowered to 0.9852, implying a change in the discount rate from 2% to 6% on an annual basis. The KL criterion equals 1.25E-05. The empirical distance measure equals 0.0535, 0.0553, 0.0564 and 0.0671 for the four sample sizes. These values are similar to those in the previous paragraph. This confirms the empirical fact that it is hard to estimate β to any precision using the dynamic properties of aggregate data on consumption and interest rates.

Third, consider changing the Taylor rule weight ψ_1 in θ_0^D to 1.23 while keeping all other parameters fixed at their original values. The KL criterion equals 0.0579. The empirical distance measure equals 0.8295, 0.9883, 0.9987, 1.0000 for the four sample sizes. This suggests that it is quite feasible to differentiate between the two models with commonly used sample sizes. This also provides a sharp contrast with the first situation. There, ψ_1 and ψ_2 were simultaneously changed to more distant values, but resulted in near observational equivalence.

Finally, we search for the model in the determinacy region that is closest to that at θ_0. The resulting parameter vector is

$$\theta^D = (10, 0.999, 2.12, 1.10, 0.01, 0.57, 0.64, 0.96, 0.27, 0.49, 0.07)'.$$

The corresponding KL criterion equals 0.327. The empirical distance equals 1.00 for all the sample sizes considered. Therefore, even though the results in Section 7.1.2 show that the dynamics of the model at θ_0 can be closely replicated by other models in the indeterminacy region, this can not be done using models in the determinacy region, as the closest model in terms of KL criterion can be readily distinguished with moderate sample sizes.

7.1.5 Global identification from business cycle frequencies

This section considers two issues. First, it recomputes the empirical distance measures at the parameter values reported in Table 2 and in (23) and (24) using only business cycle frequencies (i.e., fluctuations with periods between 6 and 32 quarters). This can help quantify additional identification challenges when some frequencies are omitted. Second, for the above mentioned cases, it searches for parameter values that minimize the KL criterion constructed using only business cycle frequencies and obtains the corresponding empirical distances. This can reveal whether the model is globally identified from the business cycle frequencies alone and further, whether the resulting closest models are substantially different from those based on the full spectrum. The results are summarized below.

For the first issue, the resulting empirical distance measures are all strictly above 0.0500 and strictly increasing with the sample size. This confirms that the parameter values reported in Table 2 and in (23) and (24) are not observationally equivalent to θ_0 even after confining the analysis to business cycle frequencies. Meanwhile, the feasibility of distinguishing them from θ_0 clearly deteriorates. Specifically, the empirical distances computed at the nine parameter vectors in Table 2 equal the following (these can be contrasted with the values in the second and last rows of Table 3). For T=80: $0.0504, 0.0523, 0.0547, 0.0531, 0.0635, 0.0702, 0.0595, 0.0696, 0.1034$; for T=1000: $0.0516, 0.0592, 0.0720, 0.0621, 0.1127, 0.1518, 0.0811, 0.1345, 0.3015$. The empirical distances computed using (23) and (24) equal (these can be compared with the values in the last two columns of Table 4) 0.0540 and 0.0568 respectively for T=80, and 0.0651 and 0.0753 respectively for T=1000. Therefore, here the output growth and output gap rules get even closer to being observationally equivalent when attention is restricted to their implications for business cycle frequencies.

For the second issue we find, interestingly, that the parameter values minimizing the business-cycle-frequency-only KL are very similar to those obtained using the full spectrum. For illustration, the counterpart to the case c=0.1 in Table 2 equals:

$$(2.61, 0.996, 0.49, 0.64, 0.20, 0.87, 0.66, 0.60, 0.27, 0.57, 0.63, 0.54, -0.06, 0.26, 0.19)'.$$

The patterns of parameter changes in the rest of the cases are similar to those in Table 2. Because of the closeness in parameter values, the resulting empirical distance measures are also similar to and only slightly below those in the previous paragraph. To avoid repetition, the details are omitted.

7.2 Lubik and Schorfheide (2004)

The analysis here parallels that for An and Schorfheide (2007). As will be shown, although the two models share a similar structure, their identification properties can have important differences.

The log linearized model is

$$\begin{aligned}
y_t &= E_t y_{t+1} - \tau(r_t - E_t\pi_{t+1}) + g_t, \\
\pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - z_t), \\
r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_1 \pi_t + (1-\rho_r)\psi_2(y_t - z_t) + \varepsilon_{rt}, \\
g_t &= \rho_g g_{t-1} + \varepsilon_{gt}, \\
z_t &= \rho_z z_{t-1} + \varepsilon_{zt},
\end{aligned}$$

where y_t denotes output, π_t is inflation, r_t is the nominal interest rate, g_t is government spending and z_t captures exogenous shifts of the marginal costs of production. The shocks satisfy $\varepsilon_{rt} \sim N(0, \sigma_r^2)$, $\varepsilon_{gt} \sim N(0, \sigma_g^2)$ and $\varepsilon_{zt} \sim N(0, \sigma_z^2)$. Among the three shocks, ε_{gt} and ε_{zt} are allowed to be correlated with correlation coefficient ρ_{gz}. The structural parameters are

$$\theta^D = (\tau, \beta, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z, \rho_{gz})'.$$

The state vector is $S_t = (\pi_t, y_t, r_t, g_t, z_t, E_t\pi_{t+1}, E_t y_{t+1})'$ and the observables are r_t, y_t and π_t.

Lubik and Schorfheide (2004) applied the following transformation to the model's solutions to ensure the impulse responses are continuous at the boundary between the determinacy and indeterminacy regions: $S_t = \Theta_1 S_{t-1} + \widetilde{\Theta}_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$ with $\widetilde{\Theta}_\varepsilon = \Theta_\varepsilon + \Theta_\epsilon(\Theta_\epsilon'\Theta_\epsilon)^{-1}\Theta_\epsilon'(\Theta_\varepsilon^b - \Theta_\varepsilon)$, where Θ_1, Θ_ϵ and Θ_ε are given in this paper's Appendix A and Θ_ε^b is the counterpart of Θ_ε with ψ_1 replaced by $\widetilde{\psi}_1 = 1 - (\beta\psi_2/\kappa)(1/\beta - 1)$. We apply the same transformation in order to be consistent with their analysis. Finally, the sunspot shock ϵ_t and the sunspot parameter θ^U are specified in the same way as (17).

7.2.1 Local identification

First, consider local identification at the posterior mean reported in Lubik and Schorfheide (2004, Column 1 in Table 3):

$$\theta_0 = (\underbrace{0.69, 0.997, 0.77, 0.77, 0.17, 0.60, 0.68, 0.82, 0.23, 0.27, 1.13, 0.14}_{\theta^D}, \underbrace{-0.68, 1.74, -0.69, 0.20}_{\theta^U})'.$$

The smallest eigenvalue of $G(\theta_0)$ equals 6.2E-04, well above the default tolerance level of 1.9E-09. The absolute and relative deviation measures along the curve (7) corresponding to the smallest eigenvalue exceed E-03 after $\|\theta - \theta_0\|$ reaches 0.085. The empirical distance corresponding to this value for T=1000 equals 0.0585. This further confirms that the parameter vector is locally identified at θ_0.

Next, we draw parameter values from the posterior distribution. The following parameter bounds are imposed: $\tau \in [0.1, 1]$, $\beta \in [0.9, 0.999]$, $\kappa \in [0.01, 5]$, $\psi_1 \in [0.01, 0.9]$, $\psi_2 \in [0.01, 5]$, $\rho_r \in [0.1, 0.99]$, $\rho_g \in [0.1, 0.99]$, $\rho_z \in [0.1, 0.99]$, $\rho_{gz} \in [-0.9, 0.9]$, $\sigma_r \in [0.01, 3]$, $\sigma_g \in [0.01, 3]$, $\sigma_z \in [0.01, 3]$, $M_{r\epsilon} \in [-3, 3]$, $M_{g\epsilon} \in [-3, 3]$, $M_{z\epsilon} \in [-3, 3]$, $\sigma_\epsilon \in [0.01, 3]$. Out of 4000 draws, the smallest eigenvalues are above the default tolerance level for 3996 cases. For the remaining 4 cases, the deviation measures increase noticeably along the curve (7), with absolute deviations exceeding E-03 after $\|\theta - \theta_0\|$ reaches 0.08, and relative absolute deviations reaching E-03 when $\|\theta - \theta_0\|$ reaches 0.45. The results indicate that these 4 points are also locally identified. Therefore, the local identification property is not confined to the posterior mean, but rather is a generic feature.

Then, consider local identification properties under determinacy. The smallest eigenvalue of $G(\theta_0^D)$ equals 2.5E-06 at the following posterior mean reported in Lubik and Schorfheide (2004, Table 3):

$$\theta_0^D = (0.54, 0.992, 0.58, 2.19, 0.30, 0.84, 0.83, 0.85, 0.18, 0.18, 0.64, 0.36)'.$$

The Matlab default tolerance level equals 1.1E-11. The largest absolute and relative deviations along the curve exceed E-03 when $\|\theta - \theta_0\|$ reaches 1. The corresponding empirical distance for T=1000 equals 0.0818. Thus, θ is locally identified at θ_0. We also take random draws from the posterior distribution of Lubik and Schorfheide (2004). Theorem 1 is then applied to all the resulting values. Out of 4000 draws, 3997 cases have their smallest eigenvalues above the default tolerance level. For the remaining 3 cases, the corresponding eigenvectors point consistently to the weak identification of β. After fixing its value, the eigenvalues become clearly above the tolerance level. Therefore, like in the indeterminacy case discussed above, the local identification property of θ^D is again a generic feature.

In summary, Taylor rule parameters are locally identified in this model but not in that of An and Schorfheide (2007). These results provide strong evidence that parameter identification is a system property. The identification conclusions reached from discussing a particular equation without referring to its background system are often, at best, fragile.

7.2.2 Global identification

This subsection considers global identification at θ_0 (indeterminacy) and θ_0^D (determinacy).

Under indeterminacy, the parameters minimizing the KL criterion for $c = 0.1, 0.5, 1.0$ are reported in Table 5 and the corresponding KL values and empirical distances are reported in Table 6. They show a pattern similar to the case of An and Schorfheide (2007). On one hand, globally no parameter value is found to be observationally equivalent to θ_0. On the other hand, even with $c = 1.0$, there still exist models with dynamics that are empirically hard to distinguish from those at θ_0. In all three cases, the parameter that moves the most is $M_{g\epsilon}$. We also repeat the analysis with $M_{g\epsilon}$ fixed at its original value to examine whether the identification improves substantially. As shown in Panel (b) in Tables 5 and 6, when c is increased to 1.0, distinguishing between the models becomes empirically feasible. In addition, the parameters that change the most are: σ_ϵ for $c = 0.1$ and $M_{r\epsilon}$ for $c = 0.5$ and $c = 1.0$.

Under determinacy, the parameters minimizing the KL criterion are reported in Table 7 and the KL values and empirical distances in Table 8. The empirical distances are all above 0.0500 and grow with the sample size, showing that the model is globally identified at θ_0^D. However, their values are quite small, suggesting that the models are hard to distinguish empirically. For all three values of c, the parameters that shift the most are the weights ψ_2 and ψ_1. In fact, when fixing all parameters except ψ_2 and ψ_1 at their original values and repeating the minimization, the empirical distance values obtained are very similar. Therefore, the Taylor rule parameters are the main source behind the weak identification. In addition, we study the identification strength when ψ_2 is fixed at its original value. As shown in Panel (b) of Table 8, the identification improves substantially. Fixing ψ_1 instead of ψ_2 leads to a similar pattern of empirical distances.

Given that the Taylor rule parameters are locally identified under determinacy in this model but not in the model of An and Schorfheide (2007), it is useful to further examine the strength of this identification. To this end, we trace out the curve in (7) by varying only the four Taylor rule parameters and following the eigenvector corresponding to the smallest eigenvalue. Table 9 shows ten equally spaced points on the curve along each direction. In direction 1, the curve is terminated

when $\left\|\theta^D - \theta_0^D\right\|$ exceeds 1. In direction 2, it stops before ψ_2 turns negative. The table reveals two interesting features. First, the parameters are weakly identified. As shown in the last two columns in the Table, the empirical distance is only 0.1583 with T=1000 when the Euclidean distance from θ_0^D reaches 1.0. Second, along the curve, ψ_1 and ψ_2 move substantially in opposite directions, while ρ_r and σ_r change very little. This suggests that the effect of decreasing (increasing) the weight on the inflation target is largely offset by increasing (decreasing) the weight on the output gap. This implies that, within this model, a hawkish policy (i.e., with $\psi_1 = 2.4840$) can have similar behavioral implications as a more dovish policy (ψ_1 =1.4830), depending on the value of the output gap policy parameter.

We have also considered the identification of the alternative monetary rules as we have done for An and Schorfheide (2007). The results are very similar. That is, there exists an expected inflation rule that is observationally equivalent to the original rule and an output gap rule that is nearly so. This holds under both determinacy and indeterminacy. In addition, the analysis using business cycle frequencies returned results generally similar to those in Section 7.1.5. The details are omitted to save space.

The similarities and differences in the identification properties of the two models can be summarized as follows. (1) Both models are globally identified at the posterior mean under indeterminacy. (2) Under determinacy, the model of Lubik and Schorfheide (2004) is globally identified at the posterior mean while that of An and Schorfheide (2007) is not locally identified. For the latter, the Taylor rule parameters are the source of the lack of identification, while for the former the same parameters lead to near observational equivalence. (3) Both models possess parameters that are weakly identified, although those parameters can differ across models. For both models and under both determinacy and indeterminacy, fixing a small number of parameters can lead to a substantial improvement in global identification.

7.3 Smets and Wouters (2007)

The Smets and Wouters (2007) model has become the workhorse model both in academia and in central banks. Previously, Iskrev (2010), Komunjer and Ng (2011) and Tkachenko and Qu (2012) analyzed its local identification properties at the posterior mean under determinacy. They found that the model is not locally identified and that the restrictions imposed by Smets and Wouters (2007) during estimation are sufficient to guarantee local identification. However, the model's global identification properties have not been studied formally.

In this section, we consider the following issues: (1) study whether the model is globally identified at the posterior mean under determinacy, (2) disentangle parameters that are identified weakly, (3) study the feasibility of distinguishing between models with different policy rules, and (4) compute empirical distances between models with different frictions. We focus on the determinacy case as this has been the model's intended application. As demonstrated in the previous two subsections, similar analysis can be carried out if one instead specifies parameters values in the indeterminacy region. For completeness and ease of interpretation, the log linearized equations of the model are included in Appendix B and an annotated list of structural parameters is reported in Table 10.

The vector of observable variables includes output (y_t), consumption (c_t), investment (i_t), wage (w_t), labor hours (l_t), inflation (π_t) and the interest rate (r_t). As in Smets and Wouters (2007), five parameters are fixed as follows: $\epsilon_p = \epsilon_w = 10, \delta = 0.025, g_y = 0.18, \phi_w = 1.50$. The analysis allows the remaining 34 structural parameters to vary. They are ordered as

$$\begin{aligned} \theta^D &= (\rho_{ga}, \mu_w, \mu_p, \alpha, \psi, \varphi, \sigma_c, \lambda, \phi_p, \iota_w, \xi_w, \iota_p, \xi_p, \sigma_l, r_\pi, r_{\Delta y}, r_y, \rho, \rho_a, \rho_b, \rho_g, \\ &\quad \rho_i, \rho_r, \rho_p, \rho_w, \sigma_a, \sigma_b, \sigma_g, \sigma_i, \sigma_r, \sigma_p, \sigma_w, \overline{\gamma}, 100(1/\beta - 1))'. \end{aligned} \tag{25}$$

The bounds imposed on the parameters throughout the analysis follow those used in Smets and Wouters (2007), except the lower bound for r_π is raised to 1.1, the upper bounds for the autoregressive coefficients of the seven shocks are lowered to 0.99, and those for the moving average coefficients are lowered to 0.90: $\rho_{ga} \in [0.01, 2]$, $\mu_w \in [0.01, 0.9]$, $\mu_p \in [0.01, 0.9]$, $\alpha \in [0.01, 1]$, $\psi \in [0.01, 1]$, $\varphi \in [2, 15]$, $\sigma_c \in [0.25, 3]$, $\lambda \in [0.001, 0.99]$, $\phi_p \in [1, 3]$, $\iota_w \in [0.01, 0.99]$, $\xi_w \in [0.3, 0.95]$, $\iota_p \in [0.01, 0.99]$, $\xi_p \in [0.5, 0.95]$, $\sigma_l \in [0.25, 10]$, $r_\pi \in [1.1, 3]$, $r_{\Delta y} \in [0.01, 0.5]$, $r_y \in [0.01, 0.5]$, $\rho \in [0.5, 0.975]$, $\rho_a \in [0.01, 0.99]$, $\rho_b \in [0.01, 0.99]$, $\rho_g \in [0.01, 0.99]$, $\rho_i \in [0.01, 0.99]$, $\rho_r \in [0.01, 0.99]$, $\rho_p \in [0.01, 0.99]$, $\rho_w \in [0.01, 0.99]$, $\sigma_a \in [0.01, 3]$, $\sigma_b \in [0.025, 5]$, $\sigma_g \in [0.01, 3]$, $\sigma_i \in [0.01, 3]$, $\sigma_r \in [0.01, 3]$, $\sigma_p \in [0.01, 3]$, $\sigma_w \in [0.01, 3]$, $\overline{\gamma} \in [0.1, 0.8], 100(\beta^{-1} - 1) \in [0.01, 2])'$.

7.3.1 Global identification under determinacy

This subsection considers the global identification properties at the posterior mean obtained in Smets and Wouters (2007, Tables 1A and 1B). The latter values are also included in the first column of Table 11.

First, consider the closest models in terms of the KL criterion while excluding a progressively larger neighborhood around θ_0^D. The panel (a) in Table 11 corresponds to the cases with $c = 0.1, 0.5$ and 1.0. The corresponding KL values and empirical distances are reported in Table 12. No

observationally equivalent model is found. This is a strong result, indicating that the Smets and Wouters (2007) model is globally identified at the posterior mean after fixing the five parameters. Meanwhile, there exist models that are hard to distinguish from the original model even with $c = 1.0$. In that case, the empirical distances are 0.1159 and 0.3123 for T=150 and 1000. Interestingly, for all values of c, the closest model is always found by moving the parameter corresponding to the steady state elasticity of the capital adjustment cost function (φ).

Next, we repeat the analysis with φ fixed at 5.74. The parameters minimizing the KL criterion are reported in Panel (b) in Table 11. For $c = 0.1$ and 0.5, the improvements are relatively small with typical sample sizes. Substantial changes occur when $c = 1.0$. In addition, for all three cases, the largest change is in the elasticity of labor supply (σ_l).

The above analysis shows that the model is globally identified at the posterior mean after fixing the five parameters. This is the first time such a strong statement about this model is made. Also, it shows that among the behavioral parameters, φ and σ_l are identified weakly. The latter is consistent with the finding of Smets and Wouters (2007, p. 594) obtained from a Bayesian perspective.

7.3.2 Identification of policy rules

The original model has the following monetary policy rule

$$r_t = \rho r_{t-1} + (1-\rho)\left(r_\pi \pi_t + r_y\left(y_t - y_t^*\right)\right) + r_{\Delta_y}\left(\left(y_t - y_t^*\right) - \left(y_{t-1} - y_{t-1}^*\right)\right) + \varepsilon_t^r. \tag{26}$$

We consider replacing it by the following rule targeting expected rather than current inflation:

$$r_t = \rho r_{t-1} + (1-\rho)\left(r_\pi E(\pi_{t+1}) + r_y\left(y_t - y_t^*\right)\right) + r_{\Delta_y}\left(\left(y_t - y_t^*\right) - \left(y_{t-1} - y_{t-1}^*\right)\right) + \varepsilon_t^r.$$

The parameter vector minimizing the KL criterion equals (Here and below, the parameters are in the same order as in (25).)

$$\begin{aligned}\phi^D \;=\; & (0.52, 0.80, 0.71, 0.18, 0.58, 5.45, 1.40, 0.70, 1.60, 0.55, 0.64, 0.27, 0.67, 1.35, 2.36, 0.22, 0.08, \\ & 0.80, 0.95, 0.22, 0.97, 0.72, 0.17, 0.90, 0.96, 0.45, 0.23, 0.53, 0.45, 0.25, 0.14, 0.24, 0.45, 0.01)'.\end{aligned}$$

The resulting KL equals 0.0080.The empirical distance measures reported in Table 13 equal 0.4829 for T=150 and 0.9903 for T=1000. Therefore, unlike the previously considered small scale New Keynesian models, here there is no (near) observational equivalence between the above two policy rules.

Next, we consider two other policy rules. The first has the original output gap component removed, i.e., $r_y = 0$. The second has the output growth component shut off, i.e., $r_{\Delta y} = 0$. The parameter vectors minimizing the respective KL criterion are

$$\begin{aligned}\phi^D &= (0.52, 0.79, 0.66, 0.17, 0.51, 7.98, 1.09, 0.80, 1.56, 0.57, 0.72, 0.24, 0.64, 2.41, 1.50, 0.19, 0.00,\\ &\quad 0.72, 0.95, 0.19, 0.97, 0.68, 0.25, 0.89, 0.94, 0.46, 0.24, 0.53, 0.42, 0.24, 0.14, 0.24, 0.80, 0.01)'\end{aligned}$$

and

$$\begin{aligned}\phi^D &= (0.47, 0.84, 0.73, 0.19, 0.53, 11.51, 1.24, 0.84, 1.72, 0.58, 0.74, 0.25, 0.64, 3.23, 1.52, 0.00, 0.04,\\ &\quad 0.78, 0.96, 0.18, 0.97, 0.67, 0.09, 0.90, 0.96, 0.44, 0.25, 0.54, 0.42, 0.23, 0.14, 0.24, 0.24, 0.10)'\end{aligned}$$

with the KL values equal to 0.0499 and 0.1334, respectively. Empirical distances reported in Table 13 show that both alternative models are distinguishable from the original one at θ_0^D even in small samples. In the case of no output gap component, the empirical distance for the sample of 80 observations equals 0.7844 and reaches 0.9845 for samples of size 200. In case of shutting off the output growth component, the empirical distance already equals 0.9965 for the sample of size 80, and equals 1 for T=150 and above.

We now examine whether there exists a parameter path along which the policy rules specified as (26) are (nearly) observationally equivalent as in the model of Lubik and Schorfheide (2004). The curve in (7) is computed by changing only the five parameters in the Taylor rule. The results are summarized in Table 14. There, the empirical distance rapidly increases along the curve. For T=150, it reaches 0.9131 and 0.9510 along the two directions, where the corresponding Euclidean distances from θ_0^D are only 0.0639 and 0.0718 respectively. This pattern is in sharp contrast with that for Lubik and Schorfheide (2004). The result vividly illustrates how the identification of the policy parameters drastically changes when we move from small to medium scale DSGE models.

7.3.3 Identification of frictions

This subsection examines how the dynamics of the model are affected when the real and nominal frictions are reduced substantially from their original values. The analysis parallels that of Smets and Wouters (Section IV, Table 4) conducted from a Bayesian perspective. The relevant friction parameters are ξ_p (price stickiness), ξ_w (wage stickiness), ι_p (price indexation), ι_w (wage indexation), φ (investment adjustment cost), λ (habit persistence), ψ (elasticity of capital utilization adjustment cost), and ϕ_p (fixed costs in production). We change the parameters one at a time

to a value that effectively dampens the effect of the corresponding friction, and then search for the model closest to the one at θ_0^D in terms of the KL criterion. Table 15 contains the resulting parameter vectors, KL criteria and empirical distance measures.

First, consider the nominal frictions. Reducing the degree of nominal price and wage stickiness to 0.10 results in an empirical distance of 1.00 for T=150 in both cases. This suggests that these two frictions are very important for generating the dynamic properties of the model. A lower price stickiness leads to a substantial increase in the degree of price indexation from 0.24 to 0.99. Meanwhile, the price mark-up shock process becomes amplified. It becomes more persistent (the AR parameter changes from 0.89 to 0.97 and the MA parameter drops from 0.69 to 0.12), while its error standard deviation also increases substantially (from 0.14 to 0.34). Turning to the wage stickiness, reducing its level leads to an increase in the degree of wage indexation from 0.58 to 0.99. The wage mark-up becomes more persistent (the AR parameter changes from 0.96 to 0.98 and the MA parameter drops from 0.84 to 0.09) and its standard deviation increases substantially (from 0.24 to 0.83). Interestingly, the investment adjustment cost parameter (φ) and labor supply elasticity (σ_l) decrease substantially in both cases.

The price and wage indexation are relevant but less important, when compared with the Calvo frictions, in generating the model's dynamics. Reducing the price indexation to a very low level of 0.01 leads to an empirical distance of 0.46 for T=150. Its KL value is also the smallest among all cases that we consider. For both cases, leaving out either friction does not have any noticeable impact on the other parameters.

Now consider the four real frictions. The results are summarized below. (1) Reducing the elasticity of adjustment cost from 5.74 to 1.00 leads to an empirical distance of 1.00 for T=150. The persistence and standard deviation of the investment shock process strongly increases. The other parameters change relatively little. (2) Reducing habit formation in consumption from 0.71 to 0.10 also leads to an empirical distance of 1.00 for T=150. Here, the elasticity of capital utilization adjustment cost (ψ) increases, while the investment adjustment cost parameter (φ) and labor supply elasticity (σ_l) both decrease substantially. Among the exogenous shocks, the investment shock becomes more persistent and its standard deviation also increases. (3) Shutting down the variable capital utilization results in the smallest KL criterion among the four, with an empirical distance of 0.83 at T=150. All parameters remain roughly the same. (4) Finally, reducing the share of fixed costs in production to 10 percent leads to an empirical distance of 1.00 for T=150. The standard deviation of the technology shock increases and the process also becomes more persistent.

In summary, all four real frictions are important in generating the dynamic properties of the model with the elasticity of capital utilization adjustment cost being the least so. The conclusions reached regarding both the nominal and real frictions are consistent with the findings in Smets and Wouters (2007) obtained from a Bayesian perspective.

7.3.4 Global identification from business cycle frequencies

This subsection carries out analysis similar to that in Section 7.1.5. The findings can be summarized as follows. First, the model is still globally identified at θ_0^D when only business cycle frequencies are considered. The respective parameter values minimizing the KL defined in (13) are broadly similar to those in Table 11. These results mirror our findings for the two small scale models. Second, distinguishing between alternative policy rules is still feasible at typical sample sizes. For T=80, the distances at the minimizers of (13) are $0.1356, 0.2596, 0.8305$ and for T=150, they grow to $0.1785, 0.3743, 0.9615$ (these values can be contrasted with the second and the third rows in Table 13). This finding is quite different from the two small scale models. It is notable that the changes in empirical distances vary substantially across the policy rules. For the expected inflation and the output growth rules the empirical distances are more than halved compared to the full spectrum case. In contrast, the empirical distances for the output gap rule change relatively little.

Finally, we also repeated the analysis summarized in Table 15 using business cycle frequencies. The resulting empirical distances at the minimizers of (13) with T=80 in the same order as in the table are (these can be compared with the second row in Table 15): $0.9461, 0.9370, 0.1090, 0.2222, 0.5895, 0.7662, 0.2433, 0.6661$. Overall, the values imply that the statements made in the preceding subsection about the relative importance of various frictions also hold when confining attention to the business cycle frequencies. Also, as in the case with monetary rules, the changes in empirical distances vary substantially across frictions. The results for price and wage stickiness display very little change, while those for price and wage indexation as well as for the variable capital utilization fall dramatically to about a third of the respective values for the full spectrum. The empirical distances for the remaining real frictions decrease moderately. These findings are informative in view of the fact that the current generation of DSGE models is designed for business cycle movements, not fluctuations at very low or high frequencies.

8 Conclusion

This paper has presented a unified framework for analyzing identification in DSGE models that encompasses both determinacy and indeterminacy. In addition to providing necessary and sufficient conditions for local and global identification, it also proposes a measure of the empirical distance between DSGE models. Empirical illustrations for both small and medium scale models show that the methods are informative and simple to implement. Although the analysis has focused on the second order properties for the ease of exposition, the information from the steady states can be easily incorporated. The results obtained are also applicable to other dynamic models with well defined spectra, such as (factor augmented) vector autoregressive moving average models.

References

Altug, S. (1989): "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, 30, 889–920.

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

Arellano, M., L. P. Hansen, and E. Sentana (2012): "Underidentification?," *Journal of Econometrics*, 170, 256–280.

Benati, L., and P. Surico (2009): "VAR Analysis and the Great Moderation," *American Economic Review*, 99, 1636–1652.

Benhabib, J., and R. E. A. Farmer (1999): "Indeterminacy and Sunspots in Macroeconomics," *The Handbook of Macroeconomics*, 1, 387–448.

Benhabib, J., S. Schmitt-Grohé, and M. Uribe (2001): "Monetary Policy and Multiple Equilibria," *American Economic Review*, 91, 167–186.

Beyer, A., and R. E. A. Farmer (2004): "On the Indeterminacy of New-Keynesian Economics," European Central Bank Working Paper 323.

Blanchard, O. (1982): "Identification in Dynamic Linear Models with Rational Expectations," NBER Technical Working Paper 0024.

Boivin, J., and M. P. Giannoni (2006): "Has Monetary Policy Become More Effective?," *The Review of Economics and Statistics*, 88, 445–462.

Brillinger, D. R. (2001): *Time Series: Data Analysis and Theory*. SIAM.

Brockwell, P., and R. Davis (1991): *Time Series: Theory and Methods*. New York: Springer-Verlag, 2 edn.

Canova, F., F. Ferroni, and C. Matthes (2013): "Choosing the Variables to Estimate Singular DSGE Models," Forthcoming in Journal of Applied Econometrics.

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

Chernoff, H. (1952): "Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations," *Annals of Mathematical Statistics*, 23, 493–507.

Christiano, L. J., M. Eichenbaum, and D. Marshall (1991): "The Permanent Income Hypothesis Revisited," *Econometrica*, 59, 397–423.

Christiano, L. J., and R. J. Vigfusson (2003): "Maximum Likelihood in the Frequency Domain: the Importance of Time-to-plan," *Journal of Monetary Economics*, 50, 789–815.

Clarida, R., J. Galí, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, 115, 147–180.

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

——— (2014): "The New-Keynesian Liquidity Trap," Unpublished Manuscript.

Cox, D. R. (1961): "Tests of Separate Families of Hypotheses," in *Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability*, vol. 1, pp. 105–123. Berkeley: University of California Press.

Del Negro, M., F. X. Diebold, and F. Schorfheide (2008): "Priors from Frequency-Domain Dummy Observations," No. 310 in 2008 Meeting Papers. Society for Economic Dynamics.

Del Negro, M., and F. Schorfheide (2008): "Forming Priors for DSGE Models (and How it Affects the Assessment of Nominal Rigidities)," *Journal of Monetary Economics*, 55, 1191–1208.

Diebold, F. X., L. E. Ohanian, and J. Berkowitz (1998): "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *The Review of Economic Studies*, 65, 433–451.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2004): "Comparing Dynamic Equilibrium Models to Data: a Bayesian Approach," *Journal of Econometrics*, 123, 153–187.

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

Guerron-Quintana, P. A. (2010): "What You Match Does Matter: The Effects of Data on DSGE Estimation," *Journal of Applied Econometrics*, 25, 774–804.

Haavelmo, T. (1944): "The Probability Approach in Econometrics," *Econometrica*, 12, supplement, 1–118.

Hannan, E. J. (1970): *Multiple Time Series*. New York: John Wiley.

Hansen, L. P. (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97, 1–30.

Hansen, L. P., and T. J. Sargent (1993): "Seasonality and Approximation Errors in Rational Expectations Models," *Journal of Econometrics*, 55, 21–55.

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

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

Koopmans, T. C. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 2, 125–144.

Leeper, E. M. (1991): "Equilibria Under 'Active' and 'Passive' Monetary and Fiscal Policies," *Journal of Monetary Economics*, 27, 129–147.

Lubik, T., and F. Schorfheide (2003): "Computing Sunspot Equilibria in Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 28, 273–285.

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

Mavroeidis, S. (2010): "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review*, 100, 491–503.

Parzen, E. (1983): "Autoregressive Spectral Estimation," in *Handbook of Statistics*, ed. by D. Brillinger, and P. Krishnaiah, vol. III, pp. 221–247. Amsterdam: North Holland.

Pesaran, H. (1981): "Identification of Rational Expectations Models," *Journal of Econometrics*, 16, 375–398.

Pinsker, M. S. (1964): *Information and Information Stability of Random Variables and Processes*. San-Francisco: Holden-Day.

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.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77, 665–696.

Sims, C. A. (1993): "Rational Expectations Modeling with Seasonally Adjusted Data," *Journal of Econometrics*, 55, 9–19.

——— (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, 1123–1175.

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

Tkachenko, D., and Z. Qu (2012): "Frequency Domain Analysis of Medium Scale DSGE Models with Application to Smets and Wouters (2007)," *Advances in Econometrics*, 28, 319–385.

Wallis, K. F. (1980): "Econometric Implications of the Rational Expectations Hypothesis," *Econometrica*, 48, 49–73.

Watson, M. W. (1993): "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101, 1011–1041.

Appendix A. Model Solution under Indeterminacy

This appendix contains an alternative derivation of Lubik and Schorfheide's (2003) representation for the full set of stable solutions under indeterminacy. It also contains a normalization based on the reduced column echelon form when the degree of indeterminacy exceeds one.

The system (1) can be transformed using a QZ decomposition, i.e., there exist matrices Q, Z, Λ and Ω such that

$$\begin{aligned} Q^*\Lambda Z^* &= \Gamma_0, \\ Q^*\Omega Z^* &= \Gamma_1, \end{aligned}$$

where Q and Z are unitary, Λ and Ω are upper triangular. Let $w_t{=}Z^*S_t$ and premultiply (1) by Q:

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1\cdot} \\ Q_{2\cdot} \end{bmatrix} (\Psi\varepsilon_t + \Pi\eta_t), \tag{A.1}$$

where an ordering has been imposed such that the diagonal elements of Λ_{11} (Λ_{22}) are greater (smaller) than those of Ω_{11} (Ω_{22}) in absolute values; the matrix Q is partitioned into $Q_{1\cdot}$ and $Q_{2\cdot}$. Then, because the generalized eigenvalues corresponding to the pair Λ_{22} and Ω_{22} are unstable and ε_t and η_t are serially uncorrelated, the block of equations corresponding to $w_{2,t}$ has a stable solution if and only if $w_{2,0} = 0$ and

$$Q_{2\cdot}\Pi\eta_t = -Q_{2\cdot}\Psi\varepsilon_t \ \text{ for all } t > 0. \tag{A.2}$$

The condition (A.2) determines $Q_{2\cdot}\Pi\eta_t$ as a function of ε_t. However, it may be insufficient to determine $Q_{1\cdot}\Pi\eta_t$, in which case it will lead to indeterminacy.

Since the rows of $Q_{2\cdot}\Pi$ can be linearly dependent, Sims (2002) and Lubik and Schorfheide (2003) suggested to work with its singular value decomposition to isolate the effective restrictions imposed on η_t. This leads to

$$Q_{2\cdot}\Pi = \begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix} \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{\cdot 1}^* \\ V_{\cdot 2}^* \end{bmatrix} = U_{\cdot 1} D_{11} V_{\cdot 1}^*, \tag{A.3}$$

where $\begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix}$ and $\begin{bmatrix} V_{\cdot 1} & V_{\cdot 2} \end{bmatrix}$ are unitary matrices and D_{11} is nonsingular. The submatrices $U_{\cdot 1}$ and $V_{\cdot 1}^*$ are unique up to multiplication by a unit-phase factor $\exp(i\varphi)$ (for the real case, up to sign). The spaces spanned by $U_{\cdot 2}$ and $V_{\cdot 2}$ are also unique, although the matrices themselves are not if their column dimensions exceed one. In the latter case, as a normalization, we use the reduced column echelon form for $V_{\cdot 2}$ when implementing the relevant procedures.[4]

[4] Note that matrices with the same column space have the same unique reduced column echelon form. To see this, suppose $\widetilde{V}_{\cdot 2}$ and $\bar{V}_{\cdot 2}$ are two different choices of $V_{\cdot 2}$ in the singular value decomposition of $Q_{2\cdot}\Pi$. Because they span the same column space, they must satisfy

$$\widetilde{V}_{\cdot 2} = \bar{V}_{\cdot 2} N \tag{A.4}$$

with N being a nonsingular matrix. Let $\widetilde{V}_{\cdot 2}^R$ be the unique reduced column echelon form of $\widetilde{V}_{\cdot 2}$. Then, by definition, there exists a sequence of elementary matrices $E_1, ..., E_K$ such that $\widetilde{V}_{\cdot 2}^R = \widetilde{V}_{\cdot 2} E_1 \cdots E_K$. Thus, using (A.4), $\widetilde{V}_{\cdot 2}^R = \bar{V}_{\cdot 2} N E_1 \cdots E_K$. Further, the matrix N itself equals the product of elementary matrices because it is nonsingular. Therefore, $\widetilde{V}_{\cdot 2}^R$ is also the unique reduced column echelon form of $\bar{V}_{\cdot 2}$.

Applying (A.3), (A.2) can be equivalently represented as

$$U_{.1}D_{11}V_{.1}^{*}\eta_t = -Q_{2.}\Psi\varepsilon_t \text{ for all } t > 0. \tag{A.5}$$

Premultiplying (A.5) by the conjugate transpose of $[\ U_{.1}\ \ U_{.2}\ ]$ does not alter the restrictions because the latter is nonsingular. Thus, (A.5) is equivalent to (using $U_{.1}^{*}U_{.1} = I$ and $U_{.2}^{*}U_{.1} = 0$)

$$\begin{bmatrix} D_{11}V_{.1}^{*}\eta_t \\ 0 \end{bmatrix} = \begin{bmatrix} -U_{.1}^{*}Q_{2.}\Psi\varepsilon_t \\ -U_{.2}^{*}Q_{2.}\Psi\varepsilon_t \end{bmatrix} \text{ for all } t > 0.$$

The second block of equations places no restrictions on η_t. The first block is equivalent to

$$V_{.1}^{*}\eta_t = -D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t.$$

This can be viewed as a system of linear equations of the form $Ax = b$ with $A = V_{.1}^{*}$ and $b = -D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t$. The full set of solutions for such a system is given by

$$\{p + v : v \text{ is any solution to } Ax = 0 \text{ and } p \text{ is a specific solution to } Ax = b\}. \tag{A.6}$$

Here, a specific solution is given by $p = -V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t$ while η_t solves $V_{.1}^{*}\eta_t = 0$ if and only if $\eta_t = V_{.2}\epsilon_t$ with ϵ_t being an arbitrary vector conformable with $V_{.2}$. Therefore, the full set of solutions to (A.5) can be represented as

$$\left\{\eta_t : \eta_t = -V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t + V_{.2}\epsilon_t \text{ with } E_{t-1}\epsilon_t = 0\right\}. \tag{A.7}$$

The restriction $E_{t-1}\epsilon_t = 0$ follows because η_t is an expectation error and $E_{t-1}\varepsilon_t = 0$. This representation is the same as in Proposition 1 in Lubik and Schorfheide (2003). It allows for determinacy as a special case.

We now provide some computational details on how to use (A.7) to solve for S_t in (1) as in Sims (2002). Define a matrix Φ as the projection coefficients of the rows of $Q_{1.}\Pi$ onto those of $Q_{2.}\Pi$:

$$\Phi = Q_{1.}\Pi V_{.1}D_{11}^{-1}U_{.1}^{*}.$$

Note that $Q_{1.}\Pi - \Phi Q_{2.}\Pi = Q_{1.}\Pi - Q_{1.}\Pi V_{.1}V_{.1}^{*} = Q_{1.}\Pi\left(I - V_{.1}V_{.1}^{*}\right)$, which equals zero under determinacy. Multiplying (A.1) by

$$\begin{bmatrix} I & -\Phi \\ 0 & I \end{bmatrix}$$

leads to

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi\Lambda_{22} \\ 0 & \Lambda_{22} \end{bmatrix}\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & \Omega_{22} \end{bmatrix}\begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1.} - \Phi Q_{2.} \\ Q_{2.} \end{bmatrix}\left(\Psi\varepsilon_t + \Pi\eta_t\right).$$

Imposing the restrictions (A.2), the above system reduces to

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12}-\Phi\Lambda_{22} \\ 0 & I \end{bmatrix}\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12}-\Phi\Omega_{22} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1.}-\Phi Q_{2.} \\ 0 \end{bmatrix}(\Psi\varepsilon_t + \Pi\eta_t).$$

Further, using the expression (A.7),

$$\begin{aligned} (Q_{1.}-\Phi Q_{2.})\,\Pi\eta_t &= (Q_{1.}\Pi-\Phi Q_{2.}\Pi)\left(-V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t + V_{.2}\epsilon_t\right) \\ &= -Q_{1.}\Pi\left(I-V_{.1}V_{.1}^{*}\right)V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t + Q_{1.}\Pi\left(I-V_{.1}V_{.1}^{*}\right)V_{.2}\epsilon_t. \end{aligned}$$

The first term on the right hand side equals zero. Therefore

$$\begin{aligned} \begin{bmatrix} \Lambda_{11} & \Lambda_{12}-\Phi\Lambda_{22} \\ 0 & I \end{bmatrix}\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} &= \begin{bmatrix} \Omega_{11} & \Omega_{12}-\Phi\Omega_{22} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1.}-\Phi Q_{2.} \\ 0 \end{bmatrix}\Psi\varepsilon_t \\ &\quad + \begin{bmatrix} Q_{1.}\Pi\left(I-V_{.1}V_{.1}^{*}\right) \\ 0 \end{bmatrix}V_{.2}\epsilon_t. \end{aligned}$$

Call the most left hand side matrix G_0. Multiply both sides of the equation by ZG_0^{-1} and use $w_t = Z^{*}S_t$, then

$$S_t = \Theta_1 S_{t-1} + \Theta_\varepsilon\varepsilon_t + \Theta_\epsilon\epsilon_t,$$

where

$$\begin{aligned} \Theta_1 &= ZG_0^{-1}\begin{bmatrix} \Omega_{11} & \Omega_{12}-\Phi\Omega_{22} \\ 0 & 0 \end{bmatrix}Z^{*}, \\ \Theta_\varepsilon &= ZG_0^{-1}\begin{bmatrix} Q_{1.}-\Phi Q_{2.} \\ 0 \end{bmatrix}\Psi, \\ \Theta_\epsilon &= ZG_0^{-1}\begin{bmatrix} Q_{1.}\Pi\left(I-V_{.1}V_{.1}^{*}\right) \\ 0 \end{bmatrix}V_{.2}. \end{aligned}$$

Because

$$G_0^{-1} = \begin{bmatrix} \Lambda_{11}^{-1} & -\Lambda_{11}^{-1}\left(\Lambda_{12}-\Phi\Lambda_{22}\right) \\ 0 & I \end{bmatrix},$$

the above matrices can be equivalently represented as $\Theta_1 = Z_{.1}\Lambda_{11}^{-1}[\ \Omega_{11}\ \ \Omega_{12}-\Phi\Omega_{22}\ ]Z^{*}$, $\Theta_\varepsilon = Z_{.1}\Lambda_{11}^{-1}\left(Q_{1.}-\Phi Q_{2.}\right)\Psi$ and $\Theta_\epsilon = Z_{.1}\Lambda_{11}^{-1}Q_{1.}\Pi\left(I-V_{.1}V_{.1}^{*}\right)V_{.2}$, where $Z_{.1}$ includes the first block of columns of Z conformable with Λ_{11}.

Appendix B. The Equations and Parameters of the Smets and Wouters (2007) Model

The model has seven observable endogenous variables with seven exogenous shocks. Below is an outline of the log linearized system. They are consistent with Smets and Wouters' (2007) code.

The aggregate resource constraint: It satisfies

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon_t^g.$$

Output (y_t) is composed of consumption (c_t), investment (i_t), capital utilization costs as a function of the capital utilization rate (z_t), and exogenous spending (ε_t^g). The latter follows an AR(1) model with an i.i.d. Normal error term (η_t^g), and is also affected by the productivity shock (η_t^a) as follows:

$$\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \rho_{ga} \eta_t^a + \eta_t^g.$$

The coefficients c_y, i_y and z_y are functions of the steady state spending-output ratio (g_y), steady state output growth (γ), capital depreciation (δ), household discount factor (β), intertemporal elasticity of substitution (σ_c), fixed costs in production (ϕ_p), and share of capital in production (α): $i_y = (\gamma - 1 + \delta)k_y$, $c_y = 1 - g_y - i_y$, and $z_y = R_*^k k_y$. Here, k_y is the steady state capital-output ratio, and R_*^k is the steady state rental rate of capital: $k_y = \phi_p \left(L_*/k_*\right)^{\alpha-1} = \phi_p \left[\left((1-\alpha)/\alpha\right)\left(R_*^k/w_*\right)\right]^{\alpha-1}$ with $w_* = \left(\alpha^\alpha (1-\alpha)^{(1-\alpha)}/[\phi_p \left(R_*^k\right)^\alpha]\right)^{1/(1-\alpha)}$, and $R_*^k = \beta^{-1}\gamma^{\sigma_c} - (1-\delta)$.

Households: The consumption Euler equation is

$$c_t = c_1 c_{t-1} + (1 - c_1) E_t c_{t+1} + c_2 (l_t - E_t l_{t+1}) - c_3 (r_t - E_t \pi_{t+1}) - \varepsilon_t^b. \tag{B.1}$$

where l_t is hours worked, r_t is the nominal interest rate, and π_t is inflation. The disturbance ε_t^b follows

$$\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \eta_t^b.$$

The relationship of the coefficients in (B.1) to the habit persistence (λ), steady state labor market mark-up (ϕ_w), and other structural parameters highlighted above is

$$c_1 = \frac{\lambda/\gamma}{1 + \lambda/\gamma}, c_2 = \frac{(\sigma_c - 1)\left(w_*^h L_*/c_*\right)}{\sigma_c \left(1 + \lambda/\gamma\right)}, c_3 = \frac{1 - \lambda/\gamma}{\left(1 + \lambda/\gamma\right)\sigma_c},$$

where

$$w_*^h L_*/c_* = \frac{1}{\phi_w} \frac{1-\alpha}{\alpha} R_*^k k_y \frac{1}{c_y},$$

where R_*^k and k_y are defined as above, and $c_y = 1 - g_y - (\gamma - 1 + \delta)k_y$.

The dynamics of households' investment are given by

$$i_t = i_1 i_{t-1} + (1 - i_1) E_t i_{+1} + i_2 q_t + \varepsilon_t^i,$$

where ε_t^i is a disturbance to the investment specific technology process, given by

$$\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \eta_t^i.$$

The coefficients satisfy

$$i_1 = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}},\ i_2 = \frac{1}{\left(1+\beta\gamma^{(1-\sigma_c)}\right)\gamma^2\varphi},$$

where φ is the steady state elasticity of the capital adjustment cost function. The corresponding arbitrage equation for the value of capital is given by

$$q_t = q_1 E_t q_{t+1} + (1-q_1) E_t r_{t+1}^k - (r_t - E_t\pi_{t+1}) - \frac{1}{c_3}\varepsilon_t^b\ , \tag{B.2}$$

with $q_1 = \beta\gamma^{-\sigma_c}(1-\delta) = (1-\delta)/(R_*^k + 1 - \delta)$.

Final and intermediate goods market: The aggregate production function is

$$y_t = \phi_p\left(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a\right),$$

where α captures the share of capital in production, and the parameter ϕ_p is one plus the fixed costs in production. Total factor productivity follows the AR(1) process

$$\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \eta_t^a.$$

The current capital service usage (k_t^s) is a function of capital installed in the previous period (k_{t-1}) and the degree of capital utilization (z_t):

$$k_t^s = k_{t-1} + z_t.$$

Furthermore, the capital utilization is a positive fraction of the rental rate of capital (r_t^k):

$$z_t = z_1 r_t^k, \quad \text{where } z_1 = (1-\psi)/\psi,$$

and ψ is a positive function of the elasticity of the capital utilization adjustment cost function and normalized to be between zero and one. The accumulation of installed capital (k_t) satisfies

$$k_t = k_1 k_{t-1} + (1-k_1) i_t + k_2 \varepsilon_t^i,$$

where ε_t^i is the investment specific technology process as defined before, and k_1 and k_2 satisfy

$$k_1 = \frac{1-\delta}{\gamma},\ k_2 = \left(1 - \frac{1-\delta}{\gamma}\right)\left(1+\beta\gamma^{(1-\sigma_c)}\right)\gamma^2\varphi.$$

The price mark-up satisfies

$$\mu_t^p = \alpha\left(k_t^s - l_t\right) + \varepsilon_t^a - w_t,$$

where w_t is the real wage. The New Keynesian Phillips curve is

$$\pi_t = \pi_1 \pi_{t-1} + \pi_2 E_t \pi_{t+1} - \pi_3 \mu_t^p + \varepsilon_t^p,$$

where ε_t^p is a disturbance to the price mark-up, following the ARMA(1,1) process given by

$$\varepsilon_t^p = \rho_p \varepsilon_{t-1}^p + \eta_t^p - \mu_p \eta_{t-1}^p.$$

The MA(1) term is intended to pick up some of the high frequency fluctuations in prices. The Phillips curve coefficients depend on price indexation (ι_p) and stickiness (ξ_p), the curvature of the goods market Kimball aggregator (ε_p), and other structural parameters:

$$\pi_1 = \frac{\iota_p}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}, \pi_2 = \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}, \pi_3 = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}\frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_p\right)\left(1-\xi_p\right)}{\xi_p\left(\left(\phi_p-1\right)\varepsilon_p+1\right)}.$$

Finally, cost minimization by firms implies that the rental rate of capital satisfies

$$r_t^k = -\left(k_t^s - l_t\right) + w_t.$$

Labor market: The wage mark-up is

$$\mu_t^w = w_t - \left(\sigma_l l_t + \frac{1}{1-\lambda/\gamma}\left(c_t - (\lambda/\gamma)c_{t-1}\right)\right),$$

where σ_l is the elasticity of labor supply. Real wage w_t adjusts slowly according to

$$w_t = w_1 w_{t-1} + (1-w_1)\left(E_t w_{t+1} + E_t \pi_{t+1}\right) - w_2 \pi_t + w_3 \pi_{t-1} - w_4 \mu_t^w + \varepsilon_t^w,$$

where the coefficients are functions of wage indexation (ι_w) and stickiness (ξ_w) parameters, and the curvature of the labor market Kimball aggregator (ε_w):

$$\begin{aligned} w_1 &= \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}, w_2 = \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}, w_3 = \frac{\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}, \\ w_4 &= \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}\frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_w\right)\left(1-\xi_w\right)}{\xi_w\left(\left(\phi_w-1\right)\varepsilon_w+1\right)}. \end{aligned}$$

The wage mark-up disturbance follows an ARMA(1,1) process:

$$\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w.$$

Monetary policy: The empirical monetary policy reaction function is

$$r_t = \rho r_{t-1} + (1-\rho)\left(r_\pi \pi_t + r_y\left(y_t - y_t^*\right)\right) + r_{\Delta y}\left(\left(y_t - y_t^*\right) - \left(y_{t-1} - y_{t-1}^*\right)\right) + \varepsilon_t^r.$$

The monetary shock ε_t^r follows an AR(1) process:

$$\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \eta_t^r.$$

The variable y_t^* stands for the time-varying optimal output level that is the result of a flexible price-wage economy. Since the equations for the flexible price-wage economy are essentially the same as above, but with the variables μ_t^p and μ_t^w set to zero, we omit the details.

Appendix C. Proofs

The proofs for Theorem 1 and Corollary 1 are essentially the same as those for Theorem 1 and Corollary 3 in Qu and Tkachenko (2012, supplementary appendix). This is because, after the parameter augmentation, θ determines the second order properties of the process. Here we still include the full detail to make the paper self-contained.

Let

$$f_\theta(\omega)^R = \begin{bmatrix} \mathrm{Re}(f_\theta(\omega)) & \mathrm{Im}(f_\theta(\omega)) \\ -\mathrm{Im}(f_\theta(\omega)) & \mathrm{Re}(f_\theta(\omega)) \end{bmatrix}, \tag{C.1}$$

where Re() and Im() denote the real and the imaginary parts of a complex matrix, i.e., if $C = A+Bi$, then $\mathrm{Re}(C) = A$ and $\mathrm{Im}(C) = B$. Because $f_\theta(\omega)$ is Hermitian, $f_\theta(\omega)^R$ is real and symmetric (see Lemma 3.7.1(v) in Brillinger, 2001). Further, let

$$R(\omega;\theta) = \mathrm{vec}(f_\theta(\omega)^R).$$

Because the correspondence between $f_\theta(\omega)$ and $R(\omega;\theta)$ is one to one, to prove the results it suffices to consider $R(\omega;\theta)$. In addition, Lemma A1 in Qu and Tkachenko (2012, supplementary material) states that

$$\left(\frac{\partial \,\mathrm{vec}\, f_\theta(\omega)}{\partial \theta'}\right)^* \left(\frac{\partial \,\mathrm{vec}\, f_\theta(\omega)}{\partial \theta'}\right) = \frac{1}{2}\left(\frac{\partial R(\omega;\theta)}{\partial \theta'}\right)' \left(\frac{\partial R(\omega;\theta)}{\partial \theta'}\right). \tag{C.2}$$

This implies that the left hand side, therefore $G(\theta)$, is real, symmetric and positive semidefinite.

Proof of Theorem 1. The relationship (C.2) implies that $G(\theta)$ equals

$$\frac{1}{2}\int_{-\pi}^{\pi} \left(\frac{\partial R(\omega;\theta_0)}{\partial \theta'}\right)' \left(\frac{\partial R(\omega;\theta_0)}{\partial \theta'}\right) d\omega.$$

This allows us to adopt the arguments in Theorem 1 in Rothenberg (1971) to prove the result.

Suppose θ_0 is *not* locally identified. Then, there exists an infinite sequence of vectors $\{\theta_k\}_{k=1}^{\infty}$ approaching θ_0 such that, for each k,

$$R(\omega;\theta_0) = R(\omega;\theta_k) \text{ for all } \omega \in [-\pi,\pi]\ .$$

For an arbitrary $\omega \in [-\pi,\pi]$, by the mean value theorem and the differentiability of $f_\theta(\omega)$ in θ,

$$0 = R_j(\omega;\theta_k) - R_j(\omega;\theta_0) = \frac{\partial R_j(\omega;\widetilde{\theta}(j,\omega))}{\partial \theta'}(\theta_k - \theta_0),$$

where the subscript j denotes the j-th element of the vector and $\widetilde{\theta}(j,\omega)$ lies between θ_k and θ_0 and in general depends on both ω and j. Let

$$d_k = (\theta_k - \theta_0)/\left\|\theta_k - \theta_0\right\|,$$

then

$$\frac{\partial R_j(\omega;\widetilde{\theta}(j,\omega))}{\partial \theta'} d_k = 0 \text{ for every } k.$$

The sequence $\{d_k\}$ lies on the unit sphere and therefore it has a convergent subsequence with a limit point d (note that d does not depend on j or ω). Assume $\{d_k\}$ itself is the convergent subsequence. As $\theta_k \to \theta_0$, d_k approaches d and

$$\lim_{k\to\infty} \frac{\partial R_j(\omega;\widetilde{\theta}(j,\omega))}{\partial\theta'} d_k = \frac{\partial R_j(\omega;\theta_0)}{\partial\theta'} d = 0,$$

where the convergence result holds because $f_\theta(\omega)$ is continuously differentiable in θ. Because this holds for an arbitrary j, it holds for the full vector $R(\omega;\theta_0)$. Therefore

$$\frac{\partial R(\omega;\theta_0)}{\partial\theta'} d = 0,$$

which implies

$$d' \left(\frac{\partial R(\omega;\theta_0)}{\partial\theta'}\right)' \left(\frac{\partial R(\omega;\theta_0)}{\partial\theta'}\right) d = 0.$$

Because the above result holds for an arbitrary $\omega \in [-\pi,\pi]$, it also holds when integrating over $[-\pi,\pi]$. Thus

$$d' \left\{ \int_{-\pi}^{\pi} \left(\frac{\partial R(\omega;\theta_0)}{\partial\theta'}\right)' \left(\frac{\partial R(\omega;\theta_0)}{\partial\theta'}\right) d\omega \right\} d = 0.$$

Because $d \neq 0$, $G(\theta_0)$ is singular.

To show the converse, suppose that $G(\theta)$ has a constant reduced rank in a neighborhood of θ_0 denoted by $\delta(\theta_0)$. Then, consider the characteristic vector $c(\theta)$ associated with one of the zero roots of $G(\theta)$. Because

$$\int_{-\pi}^{\pi} \left(\frac{\partial R(\omega;\theta)}{\partial\theta'}\right)' \left(\frac{\partial R(\omega;\theta)}{\partial\theta'}\right) d\omega \times c(\theta) = 0,$$

we have

$$\int_{-\pi}^{\pi} \left(\frac{\partial R(\omega;\theta)}{\partial\theta'} c(\theta)\right)' \left(\frac{\partial R(\omega;\theta)}{\partial\theta'} c(\theta)\right) d\omega = 0.$$

Since the integrand is continuous in ω and always non-negative, we must have

$$\left(\frac{\partial R(\omega;\theta)}{\partial\theta'} c(\theta)\right)' \left(\frac{\partial R(\omega;\theta)}{\partial\theta'} c(\theta)\right) = 0$$

for all $\omega \in [-\pi,\pi]$ and all $\theta \in \delta(\theta_0)$. Furthermore, this implies

$$\frac{\partial R(\omega;\theta)}{\partial\theta'} c(\theta) = 0 \tag{C.3}$$

for all $\omega \in [-\pi,\pi]$ and all $\theta \in \delta(\theta_0)$. Because $G(\theta)$ is continuous and has a constant rank in $\delta(\theta_0)$, the vector $c(\theta)$ is continuous in $\delta(\theta_0)$. Consider the curve χ defined by the function $\theta(v)$ which solves for $0 \leq v \leq \bar{v}$ the differential equation

$$\begin{aligned} \frac{\partial\theta(v)}{\partial v} &= c(\theta), \\ \theta(0) &= \theta_0. \end{aligned}$$

Then,

$$\frac{\partial R(\omega;\theta(v))}{\partial v} = \frac{\partial R(\omega;\theta(v))}{\partial \theta(v)'}\frac{\partial \theta(v)}{\partial v} = \frac{\partial R(\omega;\theta(v))}{\partial \theta(v)'}c(\theta) = 0 \tag{C.4}$$

for all $\omega \in [-\pi, \pi]$ and $0 \leq v \leq \bar{v}$, where the last equality uses (C.3). Thus, $R(\omega;\theta)$ is constant on the curve χ. This implies that $f_\theta(\omega)$ is constant on the same curve and that θ_0 is locally unidentifiable. This completes the proof.

Proof of Corollary 1. Recall $\theta = (\theta^{D\prime}, \theta^{U\prime})'$. Suppose θ_0^D is *not* locally identified. Then, there exists an infinite sequence of vectors $\{\theta_k\}_{k=1}^{\infty}$ approaching θ_0 such that

$$R(\omega;\theta_0) = R(\omega;\theta_k) \text{ for all } \omega \in [-\pi, \pi] \text{ and each } k.$$

By the definition of the partial identification, $\left\{\theta_k^D\right\}$ can be chosen such that $\left\|\theta_k^D - \theta_0^D\right\| / \left\|\theta_k - \theta_0\right\| > \varepsilon$ with ε being some arbitrarily small positive number. The values of θ_k^U can either change or stay fixed in this sequence; no restrictions are imposed on them besides those in the preceding display. As in the proof of Theorem 1, in the limit, we have

$$\frac{\partial R(\omega;\theta_0)}{\partial \theta'}d = 0,$$

with $d^D \neq 0$ (where d^D is comprised of the elements in d that correspond to θ^D). Therefore, on one hand,

$$G(\theta_0)d = 0,$$

on the other hand, because $d^D \neq 0$ and, by definition, $\partial\theta_0^D/\partial\theta' = [I_{\dim(\theta^D)}, 0_{\dim(\theta^U)}]$, we have

$$\frac{\partial \theta_0^D}{\partial \theta'}d = d^D \neq 0,$$

which implies

$$G^a(\theta_0)d \neq 0.$$

Thus, we have identified a vector that falls into the orthogonal column space of $G(\theta_0)$ but not of $G^a(\theta_0)$. Because the former orthogonal space always includes the latter as a subspace, this result implies that $G^a(\theta_0)$ has a higher column rank than $G(\theta_0)$.

To show the converse, suppose that $G(\theta)$ and $G^a(\theta)$ have constant ranks in a neighborhood of θ_0 denoted by $\delta(\theta_0)$. Because the rank of $G(\theta)$ is lower than that of $G^a(\theta)$, there exists a vector $c(\theta)$ such that

$$G(\theta)c(\theta) = 0 \text{ but } G^a(\theta)c(\theta) \neq 0,$$

which implies for all $\omega \in [-\pi, \pi]$ and all $\theta \in \delta(\theta_0)$ (see arguments leading to (C.3))

$$\frac{\partial R(\omega;\theta)}{\partial \theta'}c(\theta) = 0,$$

but

$$\begin{bmatrix} \partial R(\omega;\theta)/\partial\theta' \\ \partial\theta^D/\partial\theta' \end{bmatrix} c(\theta) = \begin{bmatrix} 0 \\ c^D(\theta) \end{bmatrix} \neq 0,$$

where $c^D(\theta)$ denotes the elements in $c(\theta)$ that correspond to θ^D. Because $G(\theta)$ is continuous and has constant rank in $\delta(\theta_0)$, the vector $c(\theta)$ is continuous in $\delta(\theta_0)$. As in Theorem 1, consider the curve χ defined by the function $\theta(v)$ which solves for $0 \leq v \leq \bar{v}$ the differential equation

$$\frac{\partial \theta(v)}{\partial v} = c(\theta),\ \theta(0) = \theta_0.$$

On one hand, because $c^D(\theta) \neq 0$ and $c^D(\theta)$ is continuous in θ, points on this curve correspond to different θ^D. On the other hand,

$$\frac{\partial R(\omega; \theta(v))}{\partial v} = \frac{\partial R(\omega; \theta(v))}{\partial \theta(v)'} \frac{\partial \theta(v)}{\partial v} = \frac{\partial R(\omega; \theta(v))}{\partial \theta(v)'} c(\theta) = 0$$

for all $\omega \in [-\pi, \pi]$ and $0 \leq v \leq \bar{v}$, implying $f_\theta(\omega)$ is constant on the same curve. Therefore, θ_0^D is not locally partially identifiable.

Proof of Theorem 2. First, for any $\omega \in [-\pi, \pi]$, $f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)$ admits an eigenvalue decomposition and all the eigenvalues are strictly positive. This can be shown as follows. Because $f_{\theta_1}(\omega)$ and $f_{\theta_0}(\omega)$ are positive definite, they have well defined Cholesky decompositions

$$\begin{aligned} f_{\theta_1}(\omega) &= A^*(\omega) A(\omega), \\ f_{\theta_0}(\omega) &= B^*(\omega) B(\omega). \end{aligned}$$

Therefore,

$$f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega) = A(\omega)^{-1} A^*(\omega)^{-1} B^*(\omega) B(\omega).$$

Pre and post multiplying both sides by $A(\omega)$ and $A(\omega)^{-1}$ yields

$$A(\omega) f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega) A(\omega)^{-1} = \left(B(\omega) A(\omega)^{-1}\right)^* \left(B(\omega) A(\omega)^{-1}\right).$$

The right hand side is a positive definite matrix. It has a well defined eigen decomposition with strictly positive eigenvalues. Because the above pre and post multiplication does not affect the eigenvalues, it follows that $f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)$ also has an eigenvalue decomposition with strictly positive eigenvalues. This in turn implies that $f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)$ is Hermitian.

Write the eigenvalue decomposition of $f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)$ as

$$f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega) = V^*(\omega) \Lambda(\omega) V(\omega), \tag{C.5}$$

where all the eigenvalues in $\Lambda(\omega)$ are real and strictly positive. Let $\lambda_j(\omega)$ be the j-th largest eigenvalue, then

$$\begin{aligned} & \frac{1}{2} \left\{ \mathrm{tr}(f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)) - \log\det(f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)) - n_Y \right\} \\ = & \frac{1}{2} \sum_{j=1}^{n_Y} \left[\lambda_j(\omega) - \log \lambda_j(\omega) - 1 \right]. \end{aligned}$$

For arbitrary $\lambda_j(\omega) > 0$, we always have $\lambda_j(\omega) - \log \lambda_j(\omega) - 1 \geq 0$ with the equality holding if and only if $\lambda_j(\omega) = 1$. Integrating over $[-\pi, \pi]$, we obtain $\mathrm{KL}(\theta_0, \theta_1) \geq 0$ with the equality holding if and only if $\Lambda(\omega) = I$ for all $\omega \in [-\pi, \pi]$. However, using (C.5), the latter is equivalent to

$$f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega) = V^*(\omega) V(\omega) = I$$

for all $\omega \in [-\pi, \pi]$. This is further equivalent to $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$.

Proof of Theorem 3. We only establish the limiting distribution under the null hypothesis as the proof under the alternative hypothesis is similar. First, from Assumption 5 and the Lipschitz condition in Assumption 6,

$$\begin{aligned} & T^{1/2} \left(\frac{1}{2T} \sum_{j=1}^{T-1} \left\{ \log\det(h_{\phi_0}^{-1}(\omega_j) f_{\theta_0}(\omega_j)) - \mathrm{tr}(h_{\phi_0}^{-1}(\omega_j) f_{\theta_0}(\omega_j)) + n_Y \right\} + KL_{fh}(\theta_0, \phi_0) \right) \\ = \; & O\left(T^{-\beta+1/2}\right) = o(1). \end{aligned}$$

Second,

$$\begin{aligned} & \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \mathrm{tr}\left\{ \left(f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right) \left(I(\omega_j) - f_{\theta_0}(\omega_j) \right) \right\} \qquad \text{(C.6)} \\ = \; & \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \mathrm{vec}\left\{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right\}^* \mathrm{vec}\left\{ I(\omega_j) - f_{\theta_0}(\omega_j) \right\}. \end{aligned}$$

This satisfies a multivariate central limit theorem under the assumptions stated in the theorem. The limiting variance is given by

$$\begin{aligned} & \lim_{T\to\infty} \frac{1}{2T} \sum_{j=1}^{T-1} \mathrm{vec}\left\{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right\}^* \left(f_{\theta_0}(\omega_j)' \otimes f_{\theta_0}(\omega_j) \right) \mathrm{vec}\left\{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right\} \\ = \; & \frac{1}{4\pi} \int_{-\pi}^{\pi} \mathrm{tr}\left\{ \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \right\} d\omega. \end{aligned}$$

Under the alternative hypothesis, the same argument can be applied. This completes the proof.

Proof of Corollary 6. Relaxing the cumulant condition does not affect the asymptotic normality. Therefore, it suffices to verify that the asymptotic variances have the stated expressions. We only consider the null hypothesis as the proof under the alternative hypothesis is similar. Let $\phi(\omega) = f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j)$. Then, (C.6) can be written as

$$\frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \mathrm{tr}\left\{ \phi(\omega_j)(I(\omega_j) - f_{\theta_0}(\omega_j)) \right\} = \sum_{k,l=1}^{n_Y} \left\{ \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \phi_{kl}(\omega_j)(I_{lk}(\omega_j) - f_{\theta_0 lk}(\omega_j)) \right\},$$

where $\phi_{kl}(\omega_j)$ is the (k,l)-th element of $\phi(\omega_j)$ and other quantities are defined analogously. Denote the quantity in the curly brackets by $A(k,l)$. Applying the same argument as in Proposition 10.8.5

in Brockwell and Davis (1991), we have

$$A(k,l) = \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \sum_{a,b=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) \left(I_{ab}^{\widetilde{\varepsilon}}(\omega_j) - E I_{ab}^{\widetilde{\varepsilon}}(\omega_j) \right) H_{bk}^*(\omega_j) + o_p(1),$$

where $I_{ab}^{\widetilde{\varepsilon}}(\omega_j)$ denotes the (a,b)-th element of the periodogram of $\widetilde{\varepsilon}_t$ and $H_{bk}^*(\omega_j)$ is the (b,k)-th element of $H^*(\omega_j)$. Note that $H_{bk}^*(\omega_j) = \overline{H_{kb}(\omega_j)}$. The covariance between $A(k,l)$ and $A(m,n)$ then equals

$$\begin{aligned} & \frac{1}{4T} \sum_{j.h=1}^{T-1} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) H_{bk}^*(\omega_j) \operatorname{Cov}(I_{ab}^{\widetilde{\varepsilon}}(\omega_j), I_{cd}^{\widetilde{\varepsilon}}(\omega_h)) \overline{\phi_{mn}(\omega_h) H_{nc}(\omega_h) H_{dm}^*(\omega_h)} + o(1) \\ = & \frac{1}{4T} \sum_{j,h=1}^{T-1} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) H_{bk}^*(\omega_j) \operatorname{Cov}(I_{ab}^{\widetilde{\varepsilon}}(\omega_j), I_{cd}^{\widetilde{\varepsilon}}(\omega_h)) \phi_{nm}(\omega_h) H_{cn}^*(\omega_h) H_{md}(\omega_h) + o(1). \end{aligned}$$

Proposition 11.7.3 in Brockwell and Davis (1991) shows for $0 < \omega_j, \omega_h < \pi$:

$$\operatorname{Cov}\left(I_{ab}^{\widetilde{\varepsilon}}(\omega_j), I_{cd}^{\widetilde{\varepsilon}}(\omega_h) \right) = \begin{cases} \frac{1}{4\pi^2 T} \kappa_{abcd} + \frac{1}{4\pi^2} \sigma_{ac} \sigma_{db} & \text{if } \omega_j = \omega_h, \\ \frac{1}{4\pi^2 T} \kappa_{abcd} & \text{if } \omega_j \neq \omega_h, \end{cases}$$

where σ_{ac} is the covariance between the a-th and the c-th elements of $\widetilde{\varepsilon}_t$. Applying this result, the preceding summation equals

$$\begin{aligned} & \frac{1}{8\pi^2 T} \sum_{j=1}^{T-1} \left\{ \sum_{a,c=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) \sigma_{ac} H_{cn}^*(\omega_j) \sum_{b,d=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{nm}(\omega_j) H_{md}(\omega_j) \sigma_{db} H_{bk}^*(\omega_j) \right\} \\ & + \frac{1}{16\pi^2 T^2} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd} \left[\sum_{j=1}^{T-1} H_{bk}^*(\omega_j) \phi_{kl}(\omega_j) H_{la}(\omega_j) \sum_{h=1}^{T-1} H_{cn}^*(\omega_h) \phi_{nm}(\omega_h) H_{md}(\omega_h) \right] + o(1). \end{aligned}$$

The first term converges to

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} \phi_{kl}(\omega) f_{\theta_0 ln}(\omega) \phi_{nm}(\omega) f_{\theta_0 mk}(\omega) d\omega.$$

Upon taking summation over $1 \leq k, l, m, n \leq n_Y$, it equals

$$\begin{aligned} & \frac{1}{4\pi} \int_{-\pi}^{\pi} \operatorname{tr} \left\{ \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) f_{\theta_0}(\omega) \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) f_{\theta_0}(\omega) \right\} d\omega \\ = & \frac{1}{4\pi} \int_{-\pi}^{\pi} \operatorname{tr} \left\{ \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \right\} d\omega. \end{aligned}$$

Meanwhile, the second term summed over $1 \leq k, l, m, n \leq n_Y$ equals

$$\frac{1}{16\pi^2 T^2} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd} \left[\sum_{j=1}^{T-1} \sum_{k,l=1}^{n_Y} H^*_{bk}(\omega_j) \phi_{kl}(\omega_j) H_{la}(\omega_j) \sum_{h=1}^{T-1} \sum_{m,n=1}^{n_Y} H^*_{cn}(\omega_h) \phi_{nm}(\omega_h) H_{md}(\omega_h) \right]$$

$$\rightarrow \quad \frac{1}{16\pi^2} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} H^*(\omega) \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) H(\omega) d\omega \right]_{ba}$$

$$\times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} H^*(\omega) \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) H(\omega) d\omega \right]_{cd}.$$

Because $\kappa_{abcd} = \kappa_{bacd}$, the right hand side can also be expressed as

$$\frac{1}{16\pi^2} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} H^*(\omega) \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) H(\omega) d\omega \right]_{ab}$$

$$\times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} H^*(\omega) \left(f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega) \right) H(\omega) d\omega \right]_{cd}.$$

This completes the proof.

Table 1. Prior and posterior distributions in the AS (2007) model

	Prior				Pre-Volcker Posterior			Post-1982 Posterior		
	Distribution	Mean	SD		Mode	Mean	90% interval	Mode	Mean	90% interval
τ	Gamma	2.00	0.50	τ	2.34	2.51	[1.78, 3.38]	2.20	2.24	[1.45, 3.17]
r^*	Gamma	2.00	1.00	β	0.996	0.995	[0.991, 0.998]	0.996	0.995	[0.991, 0.998]
κ	Gamma	0.50	0.20	κ	0.44	0.49	[0.27, 0.77]	0.76	0.84	[0.51, 1.25]
ψ_1	Gamma	1.10	0.50	ψ_1	0.58	0.63	[0.31, 0.94]	2.26	2.32	[1.78, 2.92]
ψ_2	Gamma	0.25	0.15	ψ_2	0.15	0.23	[0.06, 0.49]	0.17	0.26	[0.07, 0.56]
ρ_r	Beta	0.50	0.20	ρ_r	0.88	0.87	[0.76, 0.96]	0.66	0.65	[0.56, 0.74]
ρ_g	Beta	0.70	0.10	ρ_g	0.67	0.66	[0.50, 0.81]	0.94	0.93	[0.90, 0.97]
ρ_z	Beta	0.70	0.10	ρ_z	0.62	0.60	[0.47, 0.73]	0.89	0.88	[0.83, 0.93]
σ_r	Inv. Gamma	0.31	0.16	σ_r	0.27	0.27	[0.24, 0.31]	0.21	0.23	[0.19, 0.28]
σ_g	Inv. Gamma	0.38	0.20	σ_g	0.64	0.58	[0.33,0.80]	0.74	0.77	[0.65, 0.91]
σ_z	Inv. Gamma	0.75	0.39	σ_z	0.56	0.62	[0.45, 0.82]	0.25	0.26	[0.21, 0.31]
$M_{r\epsilon}$	Normal	0.00	1.00	$M_{r\epsilon}$	0.55	0.53	[0.25, 0.79]	–	–	–
$M_{g\epsilon}$	Normal	0.00	1.00	$M_{g\epsilon}$	0.06	-0.06	[-0.47, 0.19]	–	–	–
$M_{z\epsilon}$	Normal	0.00	1.00	$M_{z\epsilon}$	0.33	0.26	[0.06, 0.49]	–	–	–
σ_ϵ	Inv. Gamma	0.25	0.13	σ_ϵ	0.16	0.19	[0.12, 0.27]	–	–	–

Note. The posterior distributions are based on 500000 draws using the Random Walk Metropolis algorithm with the scaled inverse negative Hessian evaluated at the posterior mode as the proposal covariance matrix. The first 100000 draws are discarded. The Inverse Gamma priors are as in Lubik and Schorfheide (2004): $p(\sigma|v,s) \propto \sigma^{-v-1}e^{-vs^2/2\sigma^2}$, where $v = 4$ and s equals 0.25, 0.3, 0.6 and 0.2 respectively.

Table 2. Parameter values minimizing the KL criterion, AS (2007) model

		(a) All parameters can vary			(b) τ fixed			(c) τ and ψ_2 fixed		
	θ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	2.51	**2.61**	**3.01**	**3.51**	2.51	2.51	2.51	2.51	2.51	2.51
β	0.995	0.995	0.996	0.994	0.994	0.995	0.999	0.999	0.999	0.999
κ	0.49	0.49	0.49	0.52	0.50	0.47	0.42	0.44	0.37	0.41
ψ_1	0.63	0.64	0.68	0.70	0.61	0.50	0.33	0.62	0.61	0.61
ψ_2	0.23	0.20	0.06	0.01	**0.33**	**0.73**	**1.23**	0.23	0.23	0.23
ρ_r	0.87	0.87	0.86	0.85	0.88	0.90	0.92	0.87	0.88	0.88
ρ_g	0.66	0.66	0.66	0.66	0.66	0.65	0.65	0.66	0.65	0.64
ρ_z	0.60	0.60	0.58	0.57	0.60	0.60	0.59	0.61	0.60	0.67
σ_r	0.27	0.27	0.27	0.27	0.27	0.28	0.28	0.27	0.27	0.27
σ_g	0.58	0.57	0.57	0.56	0.58	0.59	0.60	0.59	0.62	0.61
σ_z	0.62	0.63	0.65	0.69	0.63	0.58	0.50	**0.52**	0.33	0.21
$M_{r\epsilon}$	0.53	0.54	0.57	0.60	0.53	0.50	0.46	0.51	0.47	0.49
$M_{g\epsilon}$	-0.06	-0.06	-0.08	-0.10	-0.06	-0.05	-0.03	-0.04	-0.01	-0.01
$M_{z\epsilon}$	0.26	0.26	0.27	0.28	0.26	0.29	0.37	0.33	**0.76**	**1.26**
σ_ϵ	0.19	0.19	0.19	0.18	0.19	0.19	0.18	0.18	0.07	0.001

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$, where θ_0 contains parameter values of the default specification. All values are rounded to the second decimal place except for β and σ_ϵ in the last column. The bold value signifies the binding constraint.

Table 3. KL and empirical distances between θ_c and θ_0, AS (2007) model

	(a) All parameters can vary			(b) τ fixed			(c) τ and ψ_2 fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	5.36E-07	1.57E-05	7.57E-05	1.04E-05	1.65E-04	3.35E-04	5.52E-05	4.24E-04	1.20E-3
T=80	0.0509	0.0553	0.0621	0.0542	0.0679	0.0769	0.0608	0.0839	0.1147
T=150	0.0513	0.0574	0.0672	0.0558	0.0761	0.0900	0.0651	0.0996	0.1492
T=200	0.0515	0.0586	0.0703	0.0568	0.0811	0.0982	0.0677	0.1095	0.1719
T=1000	0.0534	0.0710	0.1041	0.0665	0.1400	0.2007	0.0951	0.2339	0.4648

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_c given in the columns of Table 2. The empirical distance measure equals $p_{ff}(\theta_0, \theta_c, 0.05, T)$, where T is specified in the last four rows of the Table.

Table 4. KL and empirical distances between monetary policy rules, AS (2007) model

	(a) Expected inflation rule		(b) Output growth rule	
	Indeterminacy	Determinacy	Indeterminacy	Determinacy
KL	3.38E-14	4.24E-14	1.94E-05	6.00E-05
T=80	0.0500	0.0500	0.0560	0.0614
T=150	0.0500	0.0500	0.0583	0.0659
T=200	0.0500	0.0500	0.0597	0.0686
T=1000	0.0500	0.0500	0.0738	0.0976

Note. Under indeterminacy, KL and the empirical distance measure are defined as $KL_{fh}(\theta_0, \phi)$ and $p_{fh}(\theta_0, \phi, 0.05, T)$ with h and ϕ being the spectral density and structural parameter vector of the alternative model and T specified in the last four rows of the Table. Under determinacy, they are defined as $KL_{fh}(\theta_0^D, \phi^D)$ and $p_{fh}(\theta_0^D, \phi^D, 0.05, T)$.

Table 5. Parameter values minimizing the KL criterion

Indeterminacy, LS (2004) model							
		(a) All parameters can vary			(b) $M_{g\epsilon}$ fixed		
	θ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	0.69	0.69	0.69	0.70	0.69	0.70	0.72
β	0.997	0.998	0.999	0.999	0.953	0.999	0.999
κ	0.77	0.77	0.77	0.77	0.78	0.85	0.93
ψ_1	0.77	0.77	0.77	0.78	0.76	0.74	0.71
ψ_2	0.17	0.17	0.16	0.13	0.18	0.31	0.47
ρ_r	0.60	0.60	0.60	0.60	0.60	0.61	0.61
ρ_g	0.68	0.68	0.68	0.68	0.68	0.66	0.64
ρ_z	0.82	0.82	0.82	0.82	0.82	0.83	0.83
σ_r	0.23	0.23	0.23	0.23	0.23	0.24	0.24
σ_g	0.27	0.27	0.29	0.31	0.27	0.29	0.32
σ_z	1.13	1.13	1.13	1.14	1.16	1.07	1.02
ρ_{gz}	0.14	0.15	0.11	0.07	0.13	0.17	0.20
$M_{r\epsilon}$	-0.68	-0.69	-0.64	-0.59	-0.66	**-1.18**	**-1.68**
$M_{g\epsilon}$	1.74	**1.84**	**1.24**	**0.74**	1.74	1.74	1.74
$M_{z\epsilon}$	-0.69	-0.70	-0.65	-0.61	-0.66	-0.79	-0.89
σ_ϵ	0.20	0.13	0.39	0.50	**0.10**	0.16	0.01

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$. All values are rounded to the second decimal place except for β. The bold value signifies the binding constraint.

Table 6. KL and empirical distances between θ_c and θ_0

Indeterminacy, LS (2004) model						
	(a) All parameters can vary			(b) $M_{g\epsilon}$ fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	4.06E-07	1.22E-05	5.41E-05	5.45E-06	1.10E-03	3.90E-03
T=80	0.0507	0.0550	0.0604	0.0520	0.1084	0.2021
T=150	0.0510	0.0569	0.0646	0.0531	0.1392	0.2926
T=200	0.0512	0.0580	0.0671	0.0538	0.1593	0.3518
T=1000	0.0529	0.0686	0.0941	0.0604	0.4219	0.8739

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_c given in the columns of Table 5. The empirical distance measure equals $p_{ff}(\theta_0, \theta_c, 0.05, T)$, where T is specified in the last four rows of the Table.

Table 7. Parameter values minimizing the KL criterion

Determinacy, LS (2004) model							
		(a) All parameters can vary			(b) ψ_2 fixed		
	θ_0^D	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	0.54	0.54	0.55	0.55	0.55	0.56	0.57
β	0.992	0.900	0.900	0.900	0.900	0.900	0.900
κ	0.58	0.62	0.62	0.62	0.63	0.65	0.67
ψ_1	2.19	**2.09**	1.71	1.23	**2.29**	**2.69**	**3.19**
ψ_2	0.30	**0.40**	**0.80**	**1.30**	0.30	0.30	0.30
ρ_r	0.84	0.84	0.84	0.84	0.84	0.86	0.88
ρ_g	0.83	0.83	0.83	0.83	0.83	0.84	0.84
ρ_z	0.85	0.85	0.85	0.85	0.85	0.85	0.85
σ_r	0.18	0.18	0.18	0.18	0.18	0.18	0.19
σ_g	0.18	0.18	0.18	0.18	0.18	0.19	0.19
σ_z	0.64	0.64	0.64	0.64	0.64	0.64	0.64
ρ_{gz}	0.36	0.36	0.36	0.36	0.35	0.32	0.29

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$. All values are rounded to the second decimal place except for β. The bold value signifies the binding constraint.

Table 8. KL and empirical distances between θ_c^D and θ_0^D,

Determinacy, LS (2004) model						
	(a) All parameters can vary			(b) ψ_2 fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	1.06E-07	1.28E-05	6.26E-05	1.05E-04	2.10E-03	6.20E-03
T=80	0.0504	0.0544	0.0602	0.0650	0.1479	0.2729
T=150	0.0506	0.0563	0.0648	0.0713	0.2017	0.4011
T=200	0.0507	0.0573	0.0675	0.0751	0.2373	0.4810
T=1000	0.0515	0.0683	0.0970	0.1178	0.6541	0.9659

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$ with θ_c^D given in the columns of Table 7. The empirical distance measure equals $p_{ff}(\theta_0^D, \theta_c^D, 0.05, T)$, where T is specified in the last four rows of the Table.

Table 9. Empirical distances by altering the Taylor rule parameters, LS(2004) model

	ψ_1	ψ_2	ρ_r	σ_r	$\left\|\theta_j^D - \theta_0^D\right\|$	KL	$T = 80$	$T = 1000$
θ_0^D	2.1900	0.3000	0.8400	0.1800	–	–	–	–
(a) Direction 1								
θ_1^D	2.1194	0.3720	0.8399	0.1800	0.1009	1.23E-06	0.0516	0.0555
θ_2^D	2.0486	0.4442	0.8398	0.1800	0.2019	5.17E-06	0.0533	0.0617
θ_3^D	1.9779	0.5163	0.8397	0.1801	0.3029	1.22E-05	0.0552	0.0688
θ_4^D	1.9072	0.5884	0.8396	0.1801	0.4039	2.28E-05	0.0572	0.0769
θ_5^D	1.8365	0.6605	0.8395	0.1801	0.5049	3.75E-05	0.0593	0.0863
θ_6^D	1.7658	0.7326	0.8394	0.1802	0.6059	5.70E-05	0.0617	0.0969
θ_7^D	1.6951	0.8048	0.8392	0.1802	0.7069	8.19E-05	0.0642	0.1092
θ_8^D	1.6244	0.8769	0.8391	0.1803	0.8079	1.13E-04	0.0670	0.1233
θ_9^D	1.5537	0.9490	0.8390	0.1803	0.9089	1.52E-04	0.0700	0.1395
θ_{10}^D	1.4830	1.0212	0.8389	0.1804	1.0099	1.98E-04	0.0733	0.1583
(b) Direction 2								
θ_1^D	2.2193	0.2701	0.8400	0.1800	0.0419	1.99E-07	0.0505	0.0520
θ_2^D	2.2487	0.2401	0.8401	0.1800	0.0839	7.83E-07	0.0511	0.0541
θ_3^D	2.2781	0.2101	0.8401	0.1800	0.1259	1.73E-06	0.0516	0.0562
θ_4^D	2.3075	0.1801	0.8402	0.1800	0.1679	3.02E-06	0.0521	0.0583
θ_5^D	2.3369	0.1501	0.8402	0.1800	0.2099	4.64E-06	0.0526	0.0604
θ_6^D	2.3664	0.1201	0.8403	0.1800	0.2519	6.56E-06	0.0531	0.0626
θ_7^D	2.3958	0.0901	0.8403	0.1800	0.2939	8.77E-06	0.0536	0.0648
θ_8^D	2.4252	0.0601	0.8404	0.1800	0.3359	1.12E-05	0.0541	0.0670
θ_9^D	2.4546	0.0302	0.8404	0.1799	0.3779	1.40E-05	0.0546	0.0692
θ_{10}^D	2.4840	0.0002	0.8405	0.1799	0.4199	1.70E-05	0.0551	0.0714

Note. θ_j^D represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the four parameters in the monetary policy rule. The curve is extended from θ_0^D along two directions. Along Direction 1, the curve is truncated when $\|\theta_j^D - \theta_0^D\|$ exceeds 1. Along Direction 2, the curve is truncated at the closest point to zero where ψ_2 is still positive. KL is defined as $KL_{ff}(\theta_0^D, \theta_j^D)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0^D, \theta_j^D, 0.05, T)$.

Table 10. Structural parameters in the SW (2007) model

ρ_{ga}	Cross-corr.: tech. and exog. spending shocks
μ_w	Wage mark-up shock MA
μ_p	Price mark-up shock MA
α	Share of capital in production
ψ	Elasticity of capital utilization adjustment cost
φ	Investment adjustment cost
σ_c	Elasticity of intertemporal substitution
λ	Habit persistence
ϕ_p	Fixed costs in production
ι_w	Wage indexation
ξ_w	Wage stickiness
ι_p	Price indexation
ξ_p	Price stickiness
σ_l	Labor supply elasticity
r_π	Taylor rule: inflation weight
$r_{\Delta y}$	Taylor rule: feedback from output gap change
r_y	Taylor rule: output gap weight
ρ	Taylor rule: interest rate smoothing
ρ_a	Productivity shock AR
ρ_b	Risk premium shock AR
ρ_g	Exogenous spending shock AR
ρ_i	Interest rate shock AR
ρ_r	Monetary policy shock AR
ρ_p	Price mark-up shock AR
ρ_w	Wage mark-up shock AR
σ_a	Productivity shock std. dev.
σ_b	Risk premium shock std. dev.
σ_g	Exogenous spending shock std. dev.
σ_i	Interest rate shock std. dev.
σ_r	Monetary policy shock std. dev.
σ_p	Price mark-up shock std. dev.
σ_w	Wage mark-up shock std. dev.
$\overline{\gamma}$	Trend growth rate: real GDP, Infl., Wages
$100(\beta^{-1}-1)$	Discount rate

Table 11. Parameter values miminizing the KL criterion, SW(2007) model

		(a) All parameters can vary			(b) φ fixed		
	θ_0^D	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
ρ_{ga}	0.52	0.52	0.52	0.52	0.52	0.52	0.52
μ_w	0.84	0.84	0.84	0.84	0.84	0.83	0.82
μ_p	0.69	0.69	0.69	0.69	0.69	0.70	0.71
α	0.19	0.19	0.19	0.19	0.19	0.19	0.20
ψ	0.54	0.54	0.54	0.53	0.54	0.56	0.59
φ	5.74	**5.84**	**6.24**	**6.74**	5.74	5.74	5.74
σ_c	1.38	1.38	1.38	1.38	1.37	1.42	1.48
λ	0.71	0.71	0.72	0.72	0.71	0.69	0.67
ϕ_p	1.60	1.60	1.61	1.61	1.60	1.61	1.63
ι_w	0.58	0.58	0.58	0.57	0.58	0.58	0.57
ξ_w	0.70	0.70	0.70	0.71	0.71	0.67	0.64
ι_p	0.24	0.24	0.24	0.24	0.24	0.24	0.25
ξ_p	0.66	0.66	0.66	0.66	0.66	0.66	0.66
σ_l	1.83	1.83	1.85	1.86	**1.93**	**1.33**	**0.83**
r_π	2.04	2.04	2.02	2.01	2.04	2.06	2.08
$r_{\Delta y}$	0.22	0.22	0.22	0.22	0.22	0.23	0.25
r_y	0.08	0.08	0.08	0.08	0.08	0.08	0.08
ρ	0.81	0.81	0.81	0.81	0.81	0.81	0.81
ρ_a	0.95	0.95	0.95	0.95	0.95	0.95	0.95
ρ_b	0.22	0.22	0.22	0.22	0.22	0.23	0.24
ρ_g	0.97	0.97	0.97	0.97	0.97	0.97	0.97
ρ_i	0.71	0.71	0.70	0.70	0.71	0.71	0.72
ρ_r	0.15	0.15	0.15	0.15	0.15	0.16	0.17
ρ_p	0.89	0.89	0.89	0.89	0.89	0.89	0.90
ρ_w	0.96	0.96	0.96	0.96	0.96	0.96	0.96
σ_a	0.45	0.45	0.45	0.45	0.45	0.45	0.44
σ_b	0.23	0.23	0.23	0.23	0.23	0.23	0.22
σ_g	0.53	0.53	0.53	0.53	0.53	0.53	0.53
σ_i	0.45	0.45	0.45	0.45	0.45	0.45	0.45
σ_r	0.24	0.24	0.24	0.24	0.24	0.24	0.25
σ_p	0.14	0.14	0.14	0.14	0.14	0.14	0.14
σ_w	0.24	0.24	0.24	0.24	0.24	0.24	0.24
$\overline{\gamma}$	0.43	0.43	0.43	0.42	0.45	0.34	0.24
$100(\beta^{-1}-1)$	0.16	0.17	0.19	0.22	0.15	0.22	0.29

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$. All values are rounded to the second decimal place. The bold value signifies the binding constraint.

Table 12. KL and empirical distances between θ_c^D and θ_0^D, SW(2007) model

	(a) All parameters can vary			(b) φ fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	8.15E-06	1.86E-04	6.66E-04	2.87E-05	1.00E-03	5.70E-03
T=80	0.0539	0.0706	0.0941	0.0574	0.1087	0.2582
T=150	0.0553	0.0796	0.1159	0.0603	0.1389	0.3816
T=200	0.0562	0.0852	0.1299	0.0621	0.1587	0.4594
T=1000	0.0646	0.1505	0.3123	0.0800	0.4151	0.9588

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$ with θ_c^D given in the columns of Table 11. The empirical distance equals $p_{ff}(\theta_0^D, \theta_c^D, 0.05, T)$, where T is specified in the last four rows of the Table.

Table 13. KL and empirical distances between monetary policy rules, SW (2007) model

	Expected inflation rule	Output growth rule ($r_y = 0$)	Output gap rule ($r_{\Delta_y} = 0$)
KL	0.0080	0.0499	0.1334
T=80	0.3267	0.7844	0.9965
T=150	0.4829	0.9513	1.0000
T=200	0.5754	0.9845	1.0000
T=1000	0.9903	1.0000	1.0000

Note. KL and the empirical distance measure are defined as $KL_{fh}(\theta_0^D, \phi^D)$ and $p_{fh}(\theta_0^D, \phi^D, 0.05, T)$ with h and ϕ^D being the spectral density and structural parameter vector of the alternative model and T specified in the last four rows of the Table.

Table 14. Empirical distances by altering the Taylor rule parameters, SW(2007) model

	r_π	$r_{\Delta y}$	r_y	ρ	σ_r	$\left\|\theta_j^D - \theta_0^D\right\|$	KL	T=80	T=150
θ_0^D	2.0400	0.2200	0.080	0.8100	0.2400	–	–	–	–
					a) Direction 1				
θ_1^D	2.0296	0.2233	0.0794	0.8065	0.2511	0.0160	0.0019	0.1644	0.2175
θ_2^D	2.0197	0.2267	0.0789	0.8029	0.2627	0.0320	0.0072	0.3562	0.5023
θ_3^D	2.0103	0.2302	0.0784	0.7993	0.2746	0.0480	0.0156	0.5723	0.7635
θ_4^D	2.0015	0.2338	0.0779	0.7956	0.2869	0.0639	0.0266	0.7513	0.9131
θ_5^D	1.9931	0.2375	0.0774	0.7919	0.2995	0.0799	0.0398	0.8699	0.9737
θ_6^D	1.9851	0.2412	0.0770	0.7882	0.3123	0.0958	0.0549	0.9368	0.9930
θ_7^D	1.9775	0.2450	0.0766	0.7845	0.3253	0.1117	0.0716	0.9707	0.9983
θ_8^D	1.9703	0.2488	0.0762	0.7807	0.3386	0.1276	0.0894	0.9867	0.9996
θ_9^D	1.9634	0.2527	0.0758	0.7769	0.3519	0.1435	0.1082	0.9940	0.9999
θ_{10}^D	1.9568	0.2565	0.0755	0.7731	0.3655	0.1593	0.1278	0.9973	1.0000
					b) Direction 2				
θ_1^D	2.0482	0.2176	0.0805	0.8126	0.2320	0.0120	0.0011	0.0925	0.1217
θ_2^D	2.0567	0.2153	0.0809	0.8152	0.2243	0.0240	0.0045	0.1620	0.2552
θ_3^D	2.0655	0.2131	0.0814	0.8177	0.2169	0.0360	0.0103	0.2654	0.4523
θ_4^D	2.0746	0.2109	0.0819	0.8202	0.2098	0.0480	0.0186	0.4018	0.6724
θ_5^D	2.0840	0.2089	0.0825	0.8226	0.2031	0.0599	0.0293	0.5583	0.8506
θ_6^D	2.0937	0.2069	0.0830	0.8249	0.1968	0.0718	0.0424	0.7115	0.9510
θ_7^D	2.1037	0.2050	0.0836	0.8272	0.1908	0.0837	0.0579	0.8370	0.9890
θ_8^D	2.1139	0.2033	0.0842	0.8295	0.1852	0.0956	0.0755	0.9219	0.9984
θ_9^D	2.1243	0.2016	0.0848	0.8317	0.1800	0.1075	0.0952	0.9688	0.9998
θ_{10}^D	2.1349	0.2000	0.0854	0.8338	0.1751	0.1193	0.1167	0.9897	1.0000

Note. θ_j^D represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the five parameters in the monetary policy rule. The curve is extended from θ_0^D in two directions until the empirical distance reaches 1 with T=150 or the parameter boundaries are reached. KL is defined as $KL_{ff}(\theta_0^D, \theta_j^D)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0^D, \theta_j^D, 0.05, T)$.

Table 15. The closest models with constrained real and nominal frictions, SW (2007) model

	θ_0^D	ξ_p=0.1	ξ_w=0.1	ι_p=0.01	ι_w=0.01	φ=1.0	λ=0.1	ψ=0.99	ϕ_p=1.1
KL	–	0.3019	0.2434	0.0078	0.0284	0.1133	0.1712	0.0281	0.0924
T=80	-	0.9998	0.9999	0.3038	0.6946	0.9922	0.9992	0.5808	0.9675
T=150	–	1.0000	1.0000	0.4580	0.8978	0.9999	1.0000	0.8322	0.9984
ρ_{ga}	0.52	0.44	0.55	0.52	0.52	0.59	0.62	0.56	0.60
μ_w	0.84	0.64	0.09	0.86	0.81	0.79	0.77	0.83	0.88
μ_p	0.69	0.12	0.61	0.41	0.76	0.64	0.70	0.67	0.62
α	0.19	0.18	0.20	0.19	0.21	0.20	0.18	0.16	0.14
ψ	0.54	0.96	0.80	0.54	0.51	0.79	1.00	—	0.46
φ	5.74	3.27	2.00	5.73	6.12	—	2.31	4.96	4.05
σ_c	1.38	1.50	2.22	1.38	1.41	1.69	2.22	1.57	1.36
λ	0.71	0.60	0.24	0.72	0.71	0.40	—	0.64	0.63
ϕ_p	1.60	1.93	1.64	1.59	1.63	1.44	1.47	1.62	—
ι_w	0.58	0.92	0.99	0.60	—	0.66	0.61	0.63	0.65
ξ_w	0.70	0.41	—	0.71	0.71	0.58	0.50	0.67	0.74
ι_p	0.24	0.99	0.29	—	0.30	0.20	0.27	0.23	0.16
ξ_p	0.66	—	0.50	0.67	0.67	0.61	0.65	0.63	0.80
σ_l	1.83	0.81	0.25	1.84	1.58	1.21	0.42	1.15	2.18
r_π	2.04	2.30	2.88	2.07	2.01	3.00	3.00	2.19	2.11
$r_{\Delta y}$	0.22	0.21	0.33	0.22	0.22	0.34	0.45	0.27	0.31
r_y	0.08	0.07	0.10	0.08	0.08	0.14	0.13	0.10	0.09
ρ	0.81	0.79	0.81	0.81	0.81	0.84	0.86	0.81	0.80
ρ_a	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.97
ρ_b	0.22	0.32	0.73	0.22	0.22	0.50	0.73	0.28	0.28
ρ_g	0.97	0.96	0.97	0.97	0.97	0.96	0.97	0.97	0.98
ρ_i	0.71	0.80	0.86	0.71	0.71	0.91	0.85	0.77	0.72
ρ_r	0.15	0.19	0.13	0.15	0.15	0.12	0.04	0.16	0.15
ρ_p	0.89	0.97	0.91	0.86	0.89	0.91	0.91	0.90	0.89
ρ_w	0.96	0.96	0.98	0.96	0.95	0.96	0.96	0.95	0.97
σ_a	0.45	0.41	0.43	0.45	0.44	0.47	0.47	0.44	0.60
σ_b	0.23	0.22	0.12	0.23	0.23	0.16	0.14	0.21	0.21
σ_g	0.53	0.58	0.54	0.53	0.53	0.52	0.53	0.56	0.47
σ_i	0.45	0.50	0.59	0.45	0.45	0.86	0.56	0.44	0.46
σ_r	0.24	0.24	0.28	0.24	0.24	0.29	0.30	0.25	0.25
σ_p	0.14	0.34	0.15	0.11	0.16	0.14	0.14	0.14	0.13
σ_w	0.24	0.34	0.83	0.25	0.21	0.26	0.27	0.25	0.24
$\overline{\gamma}$	0.43	0.10	0.10	0.41	0.25	0.10	0.10	0.10	0.80
$100(\beta^{-1}-1)$	0.16	0.01	0.14	0.10	0.55	0.01	0.01	0.01	0.01

Note. KL (the second row) and the empirical distance measure (the third and fourth row) are defined as $KL_{ff}(\theta_0^D, \theta^D)$ and $p_{ff}(\theta_0^D, \theta^D, 0.05, T)$. Each column contains a parameter vector that minimizes the KL criterion under a particular constraint on the friction.