

Global Identification in DSGE Models Allowing for Indeterminacy*

Zhongjun Qu[†]

Boston University

Denis Tkachenko[‡]

National University of Singapore

July 12, 2016

Abstract

This paper presents a framework for analyzing global identification in log linearized DSGE models that encompasses both determinacy and indeterminacy. First, it 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. This result has three features. (1) It can be applied across DSGE models with different structures. (2) It permits checking whether a subset of frequencies can deliver identification. (3) It delivers parameter values that yield observational equivalence if there is identification failure. Next, the paper proposes a measure for the empirical closeness between two DSGE models for a further understanding of the strength of identification. The measure gauges the feasibility of distinguishing one model from another based on a finite number of observations generated by the two models. It is shown to represent the highest possible power under Gaussianity when considering local alternatives. The above theory is illustrated using two small scale and one medium scale DSGE models. The results document that certain 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. For implementation, two procedures are developed and made available, both of which can be used to obtain and thus to cross validate the findings reported in the empirical applications. Although the paper focuses on DSGE models, the results are also applicable to other vector linear processes with well defined spectra, such as the (factor augmented) vector autoregression.

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 the editor Stéphane Bonhomme and three anonymous referees for their stimulating comments and suggestions. We thank Bertille Antoine, Jesús Fernández-Villaverde, Frank Kleibergen, Arthur Lewbel, Eric Renault, Yixiao Sun and seminar participants at Academia Sinica, BC, Brown, MSU, NYU, Birmingham, CIREQ, NAWM, NBER summer institute and Narodowy Bank Polski workshop for helpful discussions and suggestions. Tkachenko acknowledges support from the Singapore Ministry of Education AcRF Tier 1 (FY2015-FRC2-009).

[†]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

DSGE models provide a unified framework for analyzing business cycles, understanding monetary policy and forecasting. In such models, understanding identification 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 in these models. Iskrev (2010) gave sufficient conditions, while Komunjer and Ng (2011) and Qu and Tkachenko (2012) gave necessary and sufficient conditions for local identification. These conditions share two substantial limitations. First, they assume determinacy. Second, they are silent about global identification. The DSGE literature contains no published work that provides necessary and sufficient conditions for global identification, even under determinacy, except when the solution is a finite order vector autoregression (Rubio-Ramírez, Waggoner and Zha, 2010 and Fukač, Waggoner and Zha, 2007). This paper aims to make progress along three dimensions. First, it generalizes the local identification results in Qu and Tkachenko (2012) to allow for indeterminacy. Second, it provides necessary and sufficient conditions for global identification. Third, it proposes an empirical distance measure to gauge the feasibility of distinguishing one model from another based on finite sample sizes.

The paper begins by setting up a platform that encompasses both determinacy and indeterminacy. First, it defines an augmented parameter vector that contains both the structural and the sunspot parameters. Next, it constructs a spectral density that characterizes the dynamic properties of the entire set of solutions. Then, it considers the spectral density as an infinite dimensional mapping, translating the identification issue into one that concerns properties of this mapping under local or global perturbations of the parameters. This platform rests on the following two properties of DSGE models, that they are general equilibrium models and that their solutions are vector linear processes. The first property makes the issue of observed exogenous processes irrelevant. Together, the two properties imply that the spectral density summarizes all the relevant information regarding the model's second order properties, irrespective whether we face determinacy or indeterminacy.

The global identification analysis in DSGE models faces two challenges. (1) The solutions are often not representable as regressions with a finite number of predetermined regressors. (2) The parameters enter and interact in a highly nonlinear fashion. These make the results in Rothenberg (1971) inapplicable. This paper explores a different route. First, it introduces a frequency domain expression for the Kullback–Leibler distance between two DSGE models. This expression depends only on the spectral densities and the dimension of the observables, thus can be easily computed without simulation or any reference to any data. Second, it shows that global identification fails

if and only if this criterion function equals zero when minimized over relevant parameter values. The minimization also returns a set of parameter values that yield observational equivalence under identification failure. Fernández-Villaverde and Rubio-Ramírez (2004) are the first to consider the Kullback–Leibler distance in the context of DSGE models. The current paper is the first that makes use of it for checking global identification. This approach in general would not have worked without the general equilibrium feature of the model and the linear structure of the solutions.

The proposed conditions can be compared along three dimensions. (i) Broadly, identification analysis can involve two types of questions with increasing levels of complexity, i.e., identification within the same model structure and identification permitting different structures. In the current context, the first question asks whether there can exist a different parameter value within the same DSGE structure generating the same dynamics for the observables. The latter asks whether an alternatively specified DSGE structure (e.g., a model with a different monetary policy rule or with different shock processes) can generate the same dynamic properties as the benchmark structure. Local identification conditions, considered here and elsewhere in the literature, can only be used to answer the first question. The global condition developed here can address the second as well. This is the only condition in the DSGE literature with such a property. (ii) The computational cost associated with the global identification condition is substantially higher than for the local condition, especially when the dimension of the parameter vector is high. Nevertheless, this paper provides evidence showing that it can be effectively implemented for both small and medium scale models. (iii) The global identification condition requires the model to be nonsingular. For singular systems, it can still be applied to nonsingular subsystems, as demonstrated later in the paper.

When global identification holds, there may still exist parameter values difficult to distinguish when faced with finite sample sizes. For example, Del Negro and Schorfheide (2008) observed similar data support for a model with moderate price rigidity and trivial wage rigidity and one in which both rigidities are high. More generally, even models with different structures (e.g., different policy rules) can produce data dynamics that are quantitatively similar. 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 represents the highest possible power under Gaussianity when considering local alternatives. Second, 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. Third, it monotonically approaches one as the sample size increases if the Kullback–Leibler distance is positive. The development here

is related to Hansen (2007) who, working from a Bayesian decision making perspective, proposed to use Chernoff's (1952) measure to quantify statistical challenges for distinguishing between two models.

The above methods are applied to three DSGE models with the following findings. (1) Regarding the model of An and Schorfheide (2007), previously Qu and Tkachenko (2012) showed that the Taylor rule parameters are locally unidentified under determinacy. The current paper considers indeterminacy and finds consistently that the Taylor rule parameters are globally identified. When comparing different policy rules, the method detects an expected inflation rule that is observationally equivalent to a current inflation rule. Meanwhile, there exists an output growth rule that is nearly observationally equivalent to an output gap rule. (2) Regarding the model analyzed in Lubik and Schorfheide (2004, in the supplementary appendix), in contrast to the previous model, the paper finds that the Taylor rule parameters are globally identified under both determinacy and indeterminacy. This contrast suggests that identification is a system property and that conclusions reached from discussing a particular equation without referring to its background system are often, at best, fragile. Meanwhile, the exact and near observational equivalence between monetary policy rules still persists. (3) Turning to Smets and Wouters (2007), the paper finds that the model is globally identified at the posterior mean under determinacy after fixing 5 parameters as in the original paper. Different from the small scale models, the (near) observational equivalence between policy rules is no longer present. The results also show that all 8 frictions are empirically relevant in generating the model's dynamic properties. The least important ones are, among the nominal frictions, the price and wage indexation and, among the real frictions, the elasticity of capital utilization adjustment cost.

This paper expands the literature on identification in rational expectations models, where notable early contributions include Wallis (1980), Pesaran (1981) and Blanchard (1982). The 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). It is the first that takes a frequency domain perspective to study global identification in DSGE models.

The paper is structured as follows. Section 2 sets up a platform for the subsequent analysis. Section 3 considers local identification. Sections 4 and 5 study global identification. Section 6 proposes an empirical distance measure. Section 7 discusses implementation. Section 8 includes two applications. Section 9 concludes. The paper has three appendices: A contains details on

solving the model under indeterminacy following Lubik and Schorfheide (2003), B includes the proofs, while the supplementary appendix contains additional theoretical and empirical results.

2 The spectrum of a DSGE model

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

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

where S_t , ε_t and η_t contain the state variables, structural shocks and expectation errors, respectively.

The system (1) can have a continuum of stable solutions. This feature, called indeterminacy, can be valuable for understanding the sources and dynamics of macroeconomic fluctuations. Lubik and Schorfheide (2004) argued that indeterminacy is consistent with how US monetary policy was conducted between 1960:I-1979:II. Related studies include 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 and suggested a further integration of this feature into the DSGE theory. Indeterminacy constitutes a conceptual challenge for analyzing identification. Below we take three steps to overcome it. Determinacy follows as a special case.

Step 1. Model solution. Lubik and Schorfheide (2003) showed that the full set of stable solutions can be represented as $S_t = \Theta_1 S_{t-1} + \Theta_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$, where ϵ_t consists of the sunspot shocks. Appendix A gives an alternative derivation of this representation using an elementary result in matrix algebra (see (A.4) to (A.6)). There, the reduced column echelon form is used to represent the indeterminacy space when the dimension of the latter exceeds one. Determinacy corresponds to $\Theta_\epsilon = 0$.

We allow ϵ_t to be correlated with ε_t . Let $M\varepsilon_t$ be the orthogonal projection of ϵ_t onto ε_t with $M = E(\varepsilon_t \varepsilon_t') [E(\varepsilon_t \varepsilon_t')]^{-1}$ and $\tilde{\epsilon}_t$ the residuals. The following representation of ϵ_t is used throughout this paper:

$$\epsilon_t = M\varepsilon_t + \tilde{\epsilon}_t.$$

Assumption 1 Assume $E(\tilde{\epsilon}_t \tilde{\epsilon}_t') = \Sigma_\epsilon$ for all t with $\|\Sigma_\epsilon\|$ being finite. Assume $\|M\| < \infty$, $E(\varepsilon_t \tilde{\epsilon}_s') = 0$ for all t and s , and $E(\tilde{\epsilon}_t \tilde{\epsilon}_s') = 0$ for all $t \neq s$.

Step 2. Parameter augmentation. Let θ^D be a vector consisting of all the structural parameters in (1). Let $\theta^U = (\text{vec}(\Sigma_\epsilon)', \text{vec}(M)')$. Define an augmented parameter vector: $\theta = (\theta^D', \theta^U)'$.

Step 3. Computing the spectral density. Let $A(L)$ denote a matrix of finite order lag polynomials to specify the observables used for the estimation or the calibration analysis and write

$$Y_t(\theta) = A(L)S_t = H(L; \theta) \begin{pmatrix} \varepsilon_t \\ \epsilon_t \end{pmatrix}, \quad (2)$$

where $H(L; \theta) = A(L)(1 - \Theta_1 L)^{-1}[\Theta_\varepsilon, \Theta_\epsilon]$. Here, the observables are explicitly indexed by θ to signify that their dynamic properties depend on both the structural and the sunspot parameters.

Remark 1 $Y_t(\theta)$ always has a vector moving average representation. It has no vector autoregressive representation if the number of shocks in ε_t and ϵ_t exceeds the dimension of $Y_t(\theta)$. Further, the polynomial $H(L; \theta)$ can be of infinite order. Consequently, 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 unique and given by

$$f_\theta(\omega) = \frac{1}{2\pi} H(\exp(-i\omega); \theta) \Sigma(\theta) H(\exp(-i\omega); \theta)^*, \quad (3)$$

where "*" denotes the conjugate transpose and $\Sigma(\theta)$ is the covariance matrix of $(\varepsilon'_t, \epsilon'_t)'$. Let $\{Y_t\}$ denote a stochastic process whose spectral density is given by $f_{\theta_0}(\omega)$ with $\omega \in [-\pi, \pi]$.

Assumption 2 (i) $\theta^D \in \Theta^D \subset \mathbb{R}^p$ 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 $f_\theta(\omega)$ are continuous in ω and differentiable in θ ; $\|f_\theta(\omega)\| \leq C$ for all $\theta \in \Theta$ and $\omega \in [-\pi, \pi]$.

Remark 2 Influenced by Koopmans (1949, p.125), the identification analysis in econometrics often proceeds in two steps. 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 Komunjer and Ng (2011) and also in Remark 1, in DSGE models the reduced form parameters can be unidentifiable. This motivates us to adopt Haavelmo's (1944, p.99) formulation to write the joint distribution (i.e, the spectral density) directly in terms of the structural parameters. Then, we treat the elements of the spectral density as mappings from a finite dimensional parameter space to a space of complex valued functions defined over $[-\pi, \pi]$. Identification holds if and only if the overall mapping is injective. This perspective underlies both the local and the global identification analysis in this paper.

3 Local identification

The analysis here extends that of Qu and Tkachenko (2012) by building on the development in Section 2, which represents the full set of solutions as a vector linear process and fully characterizes its second order properties using the spectral density $f_\theta(\omega)$. We provide details on one result that concerns the joint identification of the structural and sunspot parameters (it generalizes Theorem 1 in Qu and Tkachenko, 2012). Several other results are included in the supplementary appendix.

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 $\theta_1 = \theta_0$.*

Assumption 3 *Define*

$$G(\theta) = \int_{-\pi}^{\pi} \left(\frac{\partial \text{vec } f_\theta(\omega)}{\partial \theta'} \right)^* \left(\frac{\partial \text{vec } f_\theta(\omega)}{\partial \theta'} \right) d\omega \quad (4)$$

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

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.*

4 Global identification

This section considers the global identification of θ at θ_0 . We start with the full spectrum and then consider the steady state and also a subset of frequencies.

Definition 2 *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 proposed condition utilizes the Kullback-Leibler distance computed in the frequency domain. To motivate the idea, suppose there is a realization from the DSGE model denoted by $\{Y_t(\theta)\}_{t=1}^T$. As $T \rightarrow \infty$, the Fourier transform at ω satisfies $(2\pi T)^{-1/2} \sum_{t=1}^T Y_t(\theta) \exp(-i\omega t) \xrightarrow{d} N_c(0, f_\theta(\omega))$, where $N_c(\cdot, \cdot)$ denotes a complex normal distribution. Evaluating this distribution at θ_0 and some θ_1 , we obtain $N_c(0, f_{\theta_0}(\omega))$ and $N_c(0, f_{\theta_1}(\omega))$ respectively. The Kullback-Leibler distance of the second distribution from the first equals $(1/2)\{\text{tr}(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - \log \det(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - n_Y\}$ with n_Y being the dimension of $Y_t(\theta)$. Averaging over $\omega \in [-\pi, \pi]$, we obtain

$$KL(\theta_0, \theta_1) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\text{tr}(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - \log \det(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - n_Y\} d\omega. \quad (5)$$

We call this the Kullback-Leibler distance between two DSGE models (and, more generally, between two vector linear processes) with spectral densities $f_{\theta_0}(\omega)$ and $f_{\theta_1}(\omega)$.¹ To compute it, the main work is in solving the model at the respective parameter values; no simulations are needed. It takes less than a minute to compute it for a medium scale model such as Smets and Wouters (2007).

The Kullback-Leibler distance is a fundamental concept in the statistics and information theory literature. The expression (5) first appeared in Pinsker (1964) as the entropy rate of one vector stationary Gaussian process with respect to another (i.e., of $Y_t(\theta_0)$ with respect to $Y_t(\theta_1)$). Later, building on his result, Parzen (1983) showed that the autoregressive spectral density estimator satisfies the maximum entropy principle. This paper is the first that utilizes the expression (5) to convert the Kullback-Leibler distance into a computational device for checking global identification.

Assumption 4 *The eigenvalues of $f_{\theta}(\omega)$ are strictly positive for all $\omega \in [-\pi, \pi]$ and all $\theta \in \Theta$.*

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 $KL(\theta_0, \theta_1) > 0$ for any $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$.*

Theorem 2 reduces the problem of checking global identification to minimizing a deterministic function. Specifically, suppose the condition in Section 3 shows that θ_0 is locally identified. Then, to study global identification, we can proceed to check whether

$$\inf_{\theta_1 \in \Theta \setminus B(\theta_0)} KL(\theta_0, \theta_1) > 0, \quad (6)$$

where $B(\theta_0)$ is an open neighborhood of θ_0 . Here, $B(\theta_0)$ serves two purposes. First, it excludes parameter values that are arbitrarily close to θ_0 , at which the criterion will be arbitrarily close to zero. 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. It is important to emphasize that $KL(\theta_0, \theta)$ is a smooth deterministic function of θ . This is crucial for the computational feasibility as further discussed in Section 7.

The more recent work of Kociecki and Kolasa (2013) also explores the computational approach to global identification, but from a time domain perspective by building upon Komunjer and Ng (2011). They search for parameter values that produce the same coefficient matrices in a minimal state space representation up to some rotation. This paper remains the only one that allows for indeterminacy and different model structures and also proposes the empirical distance measure.

¹The derivation is kept informal to ease understanding. More formally, let $L(\theta)$ denote the frequency domain approximate likelihood based on $Y_t(\theta)$; c.f. (10). Then, $T^{-1}E(L(\theta_0) - L(\theta_1)) \rightarrow KL(\theta_0, \theta_1)$, where the expectation is taken under $Y_t(\theta)$ generated with $\theta = \theta_0$. Here $L(\theta)$ can also be replaced by the time domain Gaussian likelihood.

It can be interesting to further ask whether the information contained in a subset of frequencies is already sufficient for global identification, or whether information from the steady state can improve the identification. Let $\mu(\theta)$ denote the mean of $Y_t(\theta)$. Let $W(\omega)$ be an indicator function symmetric about zero with a finite number of discontinuities to select the desired frequencies. Define

$$KL^W(\theta_0, \theta_1) = \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\omega) \{ \text{tr}(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - \log \det(f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)) - n_Y \} d\omega, \quad (7)$$

$$\overline{KL}(\theta_0, \theta_1) = KL(\theta_0, \theta_1) + \frac{1}{4\pi} (\mu(\theta_0) - \mu(\theta_1))' f_{\theta_1}^{-1}(0) (\mu(\theta_0) - \mu(\theta_1)).$$

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

Definition 3 *The parameter vector θ is said to be globally identifiable from the first and second order properties of $\{Y_t\}$ at a point θ_0 if, within Θ , $\mu(\theta_1) = \mu(\theta_0)$ and $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ necessarily imply $\theta_1 = \theta_0$.*

Corollary 2 *Let Assumptions 1, 2 and 4 hold. Assume $\mu(\theta)$ is continuously differentiable over Θ . Then, θ is globally identified at θ_0 from the first and second order properties of $\{Y_t\}$ if and only if $\overline{KL}(\theta_0, \theta_1) > 0$ for all $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$.*

When the model is singular, the above three results can be applied to its nonsingular subsystems. Let C denote a selection matrix such that $CY_t(\theta)$ is nonsingular for all $\theta \in \Theta$. Then, (5) and (7) can be constructed by replacing $f_{\theta}(\omega)$ and $\mu(\theta)$ with $Cf_{\theta}(\omega)C'$ and $C\mu(\theta)$ respectively. Such an exercise is relevant because, for singular models, likelihood functions need to be based on the nonsingular subsystems. For discussions on selecting observables in singular models, see Guerron-Quintana (2010) and Canova, Ferroni and Matthes (2014). Even for nonsingular models, such an analysis can still be useful in revealing whether considering a subsystem is sufficient for identification.

Finally, if the purpose is only to detect observational equivalence, then we can build identification conditions upon other divergence measures and metrics. Some of them do not require nonsingularity. The merit of the Kullback-Leibler divergence is that its minimizer corresponds to a model with a well understood interpretation. This feature further enables us to quantify the strength of identification using the empirical distance measure developed in Section 6.

5 Global identification allowing for different model structures

The framework can allow the DSGE models to have different structures (e.g., different monetary policy rules or determinacy properties; see Section 8). Specifically, suppose $Y_t(\theta)$ and $Z_t(\phi)$ are

two processes generated by two DSGE structures (Structures 1 and 2) with means μ_θ and ν_ϕ and 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. We treat Structure 1 at $\theta = \theta_0$ as the benchmark specification and wish to learn whether Structure 2 can generate the same dynamic properties. Define

$$KL_{fh}(\theta, \phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\text{tr}(h_\phi^{-1}(\omega)f_\theta(\omega)) - \log \det(h_\phi^{-1}(\omega)f_\theta(\omega)) - n_Y\} d\omega. \quad (8)$$

Definition 4 *We say the second order properties of Structure 2 are distinct from those of 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 3 *Let Assumptions 1, 2 and 4 hold for $Y_t(\theta)$ and $Z_t(\phi)$. Then, the second order properties of Structure 2 are distinct from those of Structure 1 at θ_0 if and only if $\inf_{\phi \in \Phi} KL_{fh}(\theta_0, \phi) > 0$.*

In applications, different model structures may involve overlapping but different sets of observables. To apply the 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, and construct $KL_{fh}(\theta, \phi)$ by replacing $f_\theta(\omega)$ and $h_\phi(\omega)$ with $C_1(\exp(-i\omega))f_\theta(\omega)C_1(\exp(-i\omega))^*$ and $C_2(\exp(-i\omega))h_\phi(\omega)C_2(\exp(-i\omega))^*$.

Analogously to Corollaries 1 and 2, the model structures can be compared using a subset of frequencies or incorporating the information from the steady state. Define

$$\begin{aligned} KL_{fh}^W(\theta, \phi) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\omega) \{\text{tr}(h_\phi^{-1}(\omega)f_\theta(\omega)) - \log \det(h_\phi^{-1}(\omega)f_\theta(\omega)) - n_Y\} d\omega, \quad (9) \\ \overline{KL}_{fh}(\theta, \phi) &= KL_{fh}(\theta, \phi) + \frac{1}{4\pi} (\mu_\theta - \nu_\phi)' h_\phi^{-1}(0) (\mu_\theta - \nu_\phi). \end{aligned}$$

Then, Structure 2 is distinct from Structure 1 at θ_0 within the frequencies selected by $W(\omega)$ if and only if $KL_{fh}^W(\theta_0, \phi) > 0$ for all $\phi \in \Phi$. Structure 2 is distinct from Structure 1 at θ_0 in the first or second order properties if $\overline{KL}_{fh}(\theta_0, \phi) > 0$ for all $\phi \in \Phi$.

The magnitudes of the six criterion functions in (5) to (9) are informative about the strength of the identification. Intuitively, if there exists a θ distant from θ_0 (or, more generally, a model structure with a very different theoretical underpinning) that makes these functions very close to zero, then in finite samples it can be nearly impossible to distinguish between these parameter values (or model structures) on the grounds of their quantitative implications. Technically, because $KL_{fh}(\theta_0, \phi_0)$ is approximately the mean of the log likelihood ratio divided by the sample size (c.f. footnote 1) and DSGE models are usually applied with fewer than 300 quarterly observations, the value of $KL_{fh}(\theta_0, \phi_0)$ being below 1E-04 is often a signal that the location of the log likelihood ratio distribution is hardly distinguishable from 0 after taking into account its positive dispersion.

This suggests that 1E-04 may be viewed as a rule of thumb indication of weak identification. This is consistent with the findings in Section 8, where the empirical distances are below 0.1200 at T=1000 when the criterion functions are below 1E-04. The next section provides the formal analysis.

6 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 spectral density $f_{\theta_0}(\omega)$ using a hypothetical sample of T observations. It enables us to interpret identification not merely as a zero-one phenomenon, but also relative to the sample sizes encountered in empirical applications.

6.1 Deriving the measure

For the purpose of motivating the measure, we consider the frequency domain approximation to the Gaussian log likelihood. Applied to the two model structures, they equal (up to constant additions)

$$\begin{aligned} L_f(\theta_0) &= -\frac{1}{2} \sum_{j=1}^{T-1} \left\{ \log \det(f_{\theta_0}(\omega_j)) + \text{tr}(f_{\theta_0}^{-1}(\omega_j)I(\omega_j)) \right\}, \\ L_h(\phi_0) &= -\frac{1}{2} \sum_{j=1}^{T-1} \left\{ \log \det(h_{\phi_0}(\omega_j)) + \text{tr}(h_{\phi_0}^{-1}(\omega_j)I(\omega_j)) \right\}, \end{aligned} \quad (10)$$

where $\omega_j = 2\pi j/T$ and $I(\omega_j) = w(\omega_j)w(\omega_j)^*$ with $w(\omega_j) = (2\pi T)^{-1/2} \sum_{t=1}^T Y_t \exp(-i\omega_j t)$. The log likelihood ratio for testing $f_{\theta_0}(\omega)$ against $h_{\phi_0}(\omega)$, after division by T , can be decomposed as (multiplying the likelihoods by 2 has no effect on the measure)

$$\begin{aligned} T^{-1}(L_h(\phi_0) - L_f(\theta_0)) &= (2T)^{-1} \sum_{j=1}^{T-1} \left\{ \log \det(h_{\phi_0}^{-1}(\omega_j)f_{\theta_0}(\omega_j)) - \text{tr}(h_{\phi_0}^{-1}(\omega_j)f_{\theta_0}(\omega_j)) + n_Y \right\} \\ &\quad + (2T)^{-1} \sum_{j=1}^{T-1} \text{tr} \left\{ \left(f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right) (I(\omega_j) - f_{\theta_0}(\omega_j)) \right\}. \end{aligned} \quad (11)$$

The first term converges to $-KL_{fh}(\theta_0, \phi_0)$. The second term, after multiplication by $T^{1/2}$, satisfies a central limit theorem when Y_t has the spectral density $f_{\theta_0}(\omega)$. A similar decomposition can be applied when Y_t has the spectral density $h_{\phi_0}(\omega)$. We now state these results formally.

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 . Let $\tilde{\varepsilon}_{ta}$ denote the a -th element of $\tilde{\varepsilon}_t = (\varepsilon'_t, \varepsilon'_t)$. Assume the joint fourth cumulant satisfies $\text{cum}(\tilde{\varepsilon}_{ta}, \tilde{\varepsilon}_{sb}, \tilde{\varepsilon}_{uc}, \tilde{\varepsilon}_{vd}) = 0$ for all $1 \leq a, b, c, d \leq \dim(\tilde{\varepsilon}_t)$ and $-\infty < t, s, u, v < \infty$.*

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:*

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

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

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

where $KL_{fh}(\theta_0, \phi_0)$ is given in (8), $KL_{hf}(\phi_0, \theta_0)$ is as in (8) but with h_ϕ and f_θ reversed,

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

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 $T^{-1/2} (L_h(\phi_0) - L_f(\theta_0))$ are readily computable from the spectral densities. The theorem therefore permits straightforward calculation of the approximate power of the likelihood ratio test of $f_{\theta_0}(\omega)$ against $h_{\phi_0}(\omega)$. To proceed, select a significance level α . Then, the first result implies that the critical value for the test equals, up to $o(1)$,

$$q_\alpha = -T^{1/2} KL_{fh}(\theta_0, \phi_0) + \sqrt{V_{fh}(\theta_0, \phi_0)} z_{1-\alpha}, \quad (12)$$

where $z_{1-\alpha}$ is the 100(1 - α)th percentile of the $N(0, 1)$ distribution. The second result implies that the testing power, up to $o(1)$, equals

$$p_{fh}(\theta_0, \phi_0, \alpha, T) = \Pr \left(Z > \frac{q_\alpha - T^{1/2} KL_{hf}(\phi_0, \theta_0)}{\sqrt{V_{hf}(\phi_0, \theta_0)}} \right), \text{ where } Z \sim N(0, 1). \quad (13)$$

We call $p_{fh}(\theta_0, \phi_0, \alpha, T)$ the empirical distance of the model with the spectral density $h_{\phi_0}(\omega)$ from the one with $f_{\theta_0}(\omega)$. It has the following features. (1) For any T and α , its value is always between 0 and 1. A higher value signifies that it is easier to distinguish between the two models. (2) If $KL_{fh}(\theta_0, \phi_0) > 0$, then the measure strictly increases with T and approaches 1 as $T \rightarrow \infty$. As with the KL criterion, the measure can be applied to structures with overlapping but different sets of observables. This permits measuring the distance between small and medium scale models.

Remark 3 *To obtain the measure, 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. Importantly, there is no need to actually write down any likelihood. The latter is only used here to derive this measure. For example, it takes about two minutes to compute the four measures in a column of Table 8 (i.e., for a medium scale model) on a 2.4GHz processor using a single core.

Here, the "empirical distance" is used in the same sense as the "Kullback-Leibler distance" and it should not be interpreted as a metric. In general, it does not satisfy the triangle inequality. When applying it, we need to be explicit about the null and the alternative model. In practice, the null model can be a benchmark that one is familiar with and is considering using for understanding business cycle fluctuations or forecasting. Currently, such a role is sometimes played by the model in Lubik and Schorfheide (2004) among the small scale models and Smets and Wouters (2007) among the medium scale models. An alternative model can arise when: (a) asking a question regarding the null model or (b) evaluating the quantitative implications of some theory or hypothesis. Some of such situations are illustrated in Section 8. There, as an example of (a), we ask (Subsection 8.2.3) whether the eight frictions are essential for producing the dynamics of the original Smets and Wouters (2007) model. There, the original model is taken to be the null and the models with the respective frictions reduced are taken to be the alternatives. As an example of (b), we evaluate the hypothesis (Subsection 8.1.3) that there can exist alternative monetary policy rules leading to exact or near observational equivalence. Finally, in situations where the classification into the null and the alternative model is unclear, it can be interesting to compute the following symmetrized measure, whose relationship with the original measure is analogous to that between the Jeffreys divergence and the Kullback-Leibler divergence: $\tilde{p}_{fh}(\theta_0, \phi_0, \alpha, T) = (1/2)[p_{fh}(\theta_0, \phi_0, \alpha, T) + p_{hf}(\phi_0, \theta_0, \alpha, T)]$.

6.2 Local asymptotic properties

The derivation has kept the values of θ_0 and ϕ_0 fixed, to encompass the comparison of different model structures and also the situation where the parameter values are far apart but the dynamic properties are potentially similar. This subsection obtains additional insights by considering two drifting sequences. The first sequence maintains the same model structure, while the second sequence allows for different model structures.

Consider the first case. Suppose the null and alternative models have spectral densities $f_{\theta_0}(\omega)$ and $f_{\theta_T}(\omega)$ respectively. The parameter θ_0 is fixed and θ_T satisfies

$$\theta_T = \theta_0 + T^{-1/2}\delta, \tag{14}$$

where δ is a vector of finite constants. The next result shows that the empirical distance measure defined in (12)-(13) is now representable as a function of δ and the information matrix associated with the null model:

$$I(\theta_0) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \text{vec} f_{\theta_0}(\omega)}{\partial \theta'} \right)^* \left\{ f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega) \right\} \left(\frac{\partial \text{vec} f_{\theta_0}(\omega)}{\partial \theta'} \right) d\omega. \quad (15)$$

Corollary 4 *Let Assumptions 1, 2, 4 and 5 hold for both $f_{\theta_0}(\omega)$ and $f_{\theta_T}(\omega)$ with $I(\theta_0)$ being positive definite. Assume the first to third order derivatives of $f_{\theta}(\omega)$ with respect to θ are bounded on an open neighborhood of θ_0 for all $\omega \in [-\pi, \pi]$. Then, we have $p_{ff}(\theta_0, \theta_T, \alpha, T) = \Pr\left(Z > z_{1-\alpha} - \sqrt{\delta' I(\theta_0) \delta}\right) + o(1)$ when T is sufficiently large, where $Z \sim N(0, 1)$ and $z_{1-\alpha}$ is the $100(1 - \alpha)$ th percentile of Z .*

Corollary 4 is proved by considering second order Taylor approximations of $KL(\theta_0, \theta_T)$ and $V_{ff}(\theta_0, \theta_T)$ around θ_0 . The purpose of requiring three bounded derivatives is to ensure that the remainder terms are asymptotically negligible. This is a mild assumption for DSGE models.

The Corollary shows vividly how the value of the measure increases from α as δ moves away from zero. Its proof in the appendix also shows that within this local neighborhood, the asymmetry vanishes in the sense that $KL(\theta_0, \theta_T) = KL(\theta_T, \theta_0) + o(T^{-1}) = (1/2) (\theta_T - \theta_0)' I(\theta_0) (\theta_T - \theta_0) + o(T^{-1})$ and $V_{ff}(\theta_0, \theta_T) = V_{ff}(\theta_T, \theta_0) + o(T^{-1}) = (\theta_T - \theta_0)' I(\theta_0) (\theta_T - \theta_0) + o(T^{-1})$. The next result derives the local asymptotic power function of the exact time domain (Neyman-Pearson) likelihood ratio test. Specifically, the Gaussian likelihood ratio equals

$$\mathcal{LR}(\theta_0, \theta_T) = \mathcal{L}(\theta_T) - \mathcal{L}(\theta_0) \quad \text{with} \quad \mathcal{L}(\theta) = \frac{1}{2} \log \det \Omega_{\theta}^{-1} - \frac{1}{2} Y' \Omega_{\theta}^{-1} Y, \quad (16)$$

where $Y = (Y_1', \dots, Y_T')'$ and Ω_{θ} is the covariance matrix of Y implied by $f_{\theta}(\omega)$. The variables Y_t are assumed to have mean zero to match with inference using only the second order properties. Allowing for a nonzero mean is straightforward. Let $\Psi(\theta_0, \delta, \alpha)$ denote the local asymptotic power function of the time domain LR test.

Corollary 5 *Let the conditions in Corollary 4 hold. Then, $\Psi(\theta_0, \delta, \alpha) = \Pr(Z > z_{1-\alpha} - \sqrt{\delta' I(\theta_0) \delta})$.*

The Corollaries 4 and 5 jointly imply that, under Gaussianity, the empirical distance measure represents the highest asymptotic power for testing θ_0 against θ_T .

We now allow for different model structures. Suppose the null model has the spectral density $f_{\theta_0}(\omega)$. Suppose the alternative model structure has a spectral density $h_{\phi_T}(\omega)$ that is given by

$$h_{\phi_T}(\omega) = f_{\theta_0}(\omega) + T^{-1/2} \delta(\omega), \quad (17)$$

where $\delta(\omega)$ is independent of T and is an n_Y dimensional Hermitian matrix for any $\omega \in [-\pi, \pi]$. Let $L_f(\theta_0)$ and $L_h(\phi_T)$ denote the frequency domain log likelihoods defined in (10), but with $h_{\phi_0}(\omega)$ replaced by $h_{\phi_T}(\omega)$. Also, let $\mathcal{L}_f(\theta_0)$ and $\mathcal{L}_h(\phi_T)$ be the time domain counterparts, defined as

$$\mathcal{L}_f(\theta_0) = \frac{1}{2} \log \det \Omega_f^{-1} - \frac{1}{2} Y' \Omega_f^{-1} Y \quad \text{and} \quad \mathcal{L}_h(\phi_T) = \frac{1}{2} \log \det \Omega_h^{-1} - \frac{1}{2} Y' \Omega_h^{-1} Y, \quad (18)$$

where Ω_f and Ω_h are covariance matrices of Y implied by $f_{\theta_0}(\omega)$ and $h_{\phi_T}(\omega)$. Finally, let $\Psi_{fh}(\theta_0, \phi_T, \alpha)$ denote the local asymptotic power function of the time domain LR test against $h_{\phi_T}(\omega)$.

Assumption 6 *There exists $C < \infty$ such that $0 < \sup_{\omega \in [-\pi, \pi]} \|\delta(\omega)\| \leq C$. The elements of $\delta(\omega)$ and $f_{\theta_0}(\omega)$ are fourth order continuously differentiable with respect to ω .*

Corollary 6 *Let Assumptions 1, 2, 4 and 5 hold for both $f_{\theta_0}(\omega)$ and $h_{\phi_T}(\omega)$. Then, under Assumption 6, we have*

$$p_{fh}(\theta_0, \phi_T, \alpha, T) = \Pr \left(Z > z_{1-\alpha} - \sqrt{\frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega} \right) + o(1)$$

and

$$\Psi_{fh}(\theta_0, \phi_T, \alpha) = \Pr \left(Z > z_{1-\alpha} - \sqrt{\frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega} \right),$$

where $Z \sim N(0, 1)$ and $z_{1-\alpha}$ is the 100(1 - α)th percentile of Z .

The Corollary implies that the empirical distance measure continues to represent the highest asymptotic power under Gaussianity even with different model structures.

The Corollary also suggests a connection between the empirical distance measure and the total variation distance. Specifically, for two Gaussian processes with spectral densities $f_{\theta_0}(\omega)$ and $h_{\phi_T}(\omega)$, an approximate lower bound on the total variation distance between them is given by $2(\Psi_{fh}(\theta_0, \phi_T, \alpha) - \alpha)$, where α can be any value between 0 and 1. The supremum of all such lower bounds, $\sup_{\alpha \in [0, 1]} 2(\Psi_{fh}(\theta_0, \phi_T, \alpha) - \alpha)$, gives an asymptotic approximation to this total variation distance by Theorem 13.1.1 in Lehmann and Romano (2008). Corollary 6 implies $\Psi_{fh}(\theta_0, \phi_T, \alpha) = p_{fh}(\theta_0, \phi_T, \alpha, T) + o(1)$. Consequently, $\sup_{\alpha \in [0, 1]} 2(p_{fh}(\theta_0, \phi_T, \alpha, T) - \alpha)$ also constitutes an approximation to this distance.

Some implications of the above results will be examined using simulations in Subsection 8.2.3, see Remark 4 there. The results in this subsection hold for other Gaussian vector linear processes with well defined spectra. This makes them of potential value beyond DSGE models.

6.3 Subsets of frequencies and the steady state

Consider the same setting as in (10). Let $L_f^W(\theta_0)$ and $V_{fh}^W(\theta_0, \phi_0)$ equal $L_f(\theta_0)$ and $V_{fh}(\theta_0, \phi_0)$, except with the integration and summation taken only over the frequencies selected by $W(\cdot)$.

Corollary 7 *Let Assumptions 1, 2, 4 and 5 hold for both $f_\theta(\omega)$ and $h_\phi(\omega)$. If the data are generated by $f_{\theta_0}(\omega)$, then $T^{1/2} \left(T^{-1}(L_h^W(\phi_0) - L_f^W(\theta_0)) + KL_{fh}^W(\theta_0, \phi_0) \right) \xrightarrow{d} N(0, V_{fh}^W(\theta_0, \phi_0))$. If the data are generated by $h_{\phi_0}(\omega)$, then $T^{1/2} \left(T^{-1}(L_h^W(\phi_0) - L_f^W(\theta_0)) - KL_{hf}^W(\phi_0, \theta_0) \right) \xrightarrow{d} N(0, V_{hf}^W(\phi_0, \theta_0))$.*

Consequently, a measure for distinguishing models using selected frequencies can be constructed in the same way as $p_{fh}(\theta_0, \phi_0, \alpha, T)$, but with $KL_{fh}(\theta_0, \phi_0)$, $KL_{hf}(\phi_0, \theta_0)$, $V_{fh}(\theta_0, \phi_0)$, $V_{hf}(\phi_0, \theta_0)$ replaced by $KL_{fh}^W(\theta_0, \phi_0)$, $KL_{hf}^W(\phi_0, \theta_0)$, $V_{fh}^W(\theta_0, \phi_0)$, $V_{hf}^W(\phi_0, \theta_0)$. We call it $p_{fh}^W(\theta_0, \phi_0, \alpha, T)$.

Suppose the spectral densities $f_\theta(\omega)$ and $h_\phi(\omega)$ are associated with the means μ_{θ_0} and ν_{ϕ_0} respectively. Then, the frequency domain likelihoods incorporating the first order properties are given by (see Hansen and Sargent, 1993) $\bar{L}_f(\theta_0) = L_f(\theta_0) - (1/2)\{\log \det(f_{\theta_0}(0)) + \text{tr}(f_{\theta_0}^{-1}(0)I_{\theta_0}(0))\}$ and $\bar{L}_h(\phi_0) = L_h(\phi_0) - (1/2)\{\log \det(h_{\phi_0}(0)) + \text{tr}(h_{\phi_0}^{-1}(0)I_{\phi_0}(0))\}$, where $I_{\theta_0}(0) = w_{\theta_0}(0)w_{\theta_0}(0)^*$, $w_{\theta_0}(0) = (2\pi T)^{-1/2} \sum_{t=1}^T (Y_t - \mu_{\theta_0})$ and $I_{\phi_0}(0)$ is defined analogously. Let $\bar{V}_{fh}(\theta_0, \phi_0) = V_{fh}(\theta_0, \phi_0) + (\mu_{\theta_0} - \nu_{\phi_0})' h_{\phi_0}^{-1}(0) (\mu_{\theta_0} - \nu_{\phi_0}) / (2\pi)$ and $\bar{V}_{hf}(\phi_0, \theta_0) = V_{hf}(\phi_0, \theta_0) + (\mu_{\theta_0} - \nu_{\phi_0})' f_{\theta_0}^{-1}(0) (\mu_{\theta_0} - \nu_{\phi_0}) / (2\pi)$.

Corollary 8 *Let Assumptions 1, 2, 4 and 5 hold for both $f_\theta(\omega)$ and $h_\phi(\omega)$. Assume μ_{θ_0} and ν_{ϕ_0} are finite and $E(\tilde{\varepsilon}_{ta}\tilde{\varepsilon}_{sb}\tilde{\varepsilon}_{uc}) = 0$ for all $1 \leq a, b, c \leq \dim(\tilde{\varepsilon}_t)$ and $-\infty < t, s, u < \infty$. Then, under μ_{θ_0} and $f_{\theta_0}(\omega)$: $T^{1/2} \left(T^{-1} (\bar{L}_h(\phi_0) - \bar{L}_f(\theta_0)) + \bar{K}\bar{L}_{fh}(\theta_0, \phi_0) \right) \xrightarrow{d} N(0, \bar{V}_{fh}(\theta_0, \phi_0))$. Under ν_{ϕ_0} and $h_{\phi_0}(\omega)$: $T^{1/2} \left(T^{-1} (\bar{L}_h(\phi_0) - \bar{L}_f(\theta_0)) - \bar{K}\bar{L}_{hf}(\phi_0, \theta_0) \right) \xrightarrow{d} N(0, \bar{V}_{hf}(\phi_0, \theta_0))$.*

Using the above result, a measure based on both the mean and the spectrum can be constructed in the same way as $p_{fh}(\theta_0, \phi_0, \alpha, T)$ but with $KL_{fh}(\theta_0, \phi_0)$, $KL_{hf}(\phi_0, \theta_0)$, $V_{fh}(\theta_0, \phi_0)$, $V_{hf}(\phi_0, \theta_0)$ replaced by $\bar{K}\bar{L}_{fh}(\theta_0, \phi_0)$, $\bar{K}\bar{L}_{hf}(\phi_0, \theta_0)$, $\bar{V}_{fh}(\theta_0, \phi_0)$, $\bar{V}_{hf}(\phi_0, \theta_0)$. We call it $\bar{p}_{fh}(\theta_0, \phi_0, \alpha, T)$.

From a technical perspective, the above two results are immediate extensions of Theorem 3. However, from an empirical perspective, their added value can be substantial. Specifically, when the three measures are applied jointly, the results can be informative about to what extent the differences between models are driven by the business cycle frequencies as opposed to other frequencies or the steady state properties. Because DSGE models are designed for medium term economic fluctuations, not very short or long term fluctuations, such information can be valuable.

7 Implementation

This section proposes procedures for implementing the identification conditions and also addresses the numerical issues that arise. The main focus is on the global analysis as the local analysis is similar to and discussed in Qu and Tkachenko (2012). The discussion uses $KL(\theta_0, \theta)$ as the illustrating condition; the statements also apply to the other global identification conditions.

7.1 Global optimization

The minimization of $KL(\theta_0, \theta)$ over θ is a first order challenge to the suggested approach. It faces two difficulties: $KL(\theta_0, \theta)$ may have multiple local minima and the dimension of θ can be high. The first feature makes standard local searches with a single initial value unreliable, while the latter makes global searches over a fine grid infeasible. Meanwhile, the problem exhibits two desirable features, i.e., $KL(\theta_0, \theta)$ is a deterministic function of θ and it is typically infinitely differentiable with respect to it. These two features help to make global optimization possible.

We carry out the minimization in two steps. The first step conducts global searches with gradient free methods. This permits exploring wide parameter regions even under multimodality. These searches return a range of parameter values that correspond to the regions where the values of $KL(\theta_0, \theta)$ are small. The second step applies multiple local searches, using the values returned from the first step along with additional uniformly randomly generated initial values within the relevant parameter bounds. These local searches exploit the smoothness in $KL(\theta_0, \theta)$ and are efficient in locating the nearby locally optimal solution. In both steps, we employ optimizers that are parallelizable, thus maintaining computational feasibility even when the dimension of θ is high.

We develop two procedures in Matlab to carry out the empirical applications. The first procedure utilizes the genetic algorithm (*ga*) in the first step and a local solver with multiple starting points (*fmincon* inside *multistart*) in the second step. The second procedure replaces the genetic algorithm with the particle swarm algorithm (*particleswarm*) followed by the same second step as above. In both procedures, the initial values are generated by these algorithms with the seed set to the Matlab default. The algorithm specifications (e.g., population sizes and iterations) are kept fixed for small scale models and further increased for the medium scale model. The procedures both lead to the results reported in the tables of the main paper and the supplementary appendix. In practice, one should carefully check for convergence. Having two optimization procedures rather than one helps to achieve such a goal. In addition, one can specify different initial values for the optimization procedures to further examine whether the results change.

The associated computational cost, in light of the applications in Section 8, can be summarized as follows (the statements pertain to a desktop with an 8-core Intel 2.4Ghz processor). For small scale models, it takes 5-18 minutes to produce a column of the Tables in the paper and the appendix. For the medium scale model, it takes 4-10 hours to produce such a column. For the latter, the mean computation times are, respectively, 4.6 and 7.4 hours for the two procedures. Experimentations suggest that the computation time declines almost linearly with the number of cores used.

In the two procedures, the neighborhood $B(\theta_0)$ is allowed to take the following general form: $B(\theta_0) = \{\theta : \|[\theta(\text{sel}) - \theta_0(\text{sel})] ./ w(\theta_0(\text{sel}))\|_r < c\}$, where $r = 1, 2$ or ∞ , “./” gives the element-by-element division, “sel” selects the elements affecting the constraint and $w(\cdot)$ is a vector of weights. In Section 8, we first set r to ∞ , $w(\cdot)$ to a vector of ones and “sel” as selecting the full parameter vector, and then carry out the following further analysis: (1) considering $r = 1$ and 2, (2) specifying $w(\cdot)$ according to some distribution, and (3) using “sel” to group the parameters into more homogenous subsets. For all cases, we examine whether the conclusion regarding the observational equivalence remains the same and whether the parameters pinpointed as weakly identified remain consistent.

Finally, to carry out the identification analysis under a different parameterization, we only need to replace $\theta(\text{sel})$ with $g(\theta(\text{sel}))$, where $g(\cdot)$ is a vector valued function that carries out the parameter transformations. We do not modify the expressions of KL or the empirical distance measure because the likelihood value is invariant to reparameterizations.

7.2 Numerical issues

Three numerical aspects reside within the above optimization problem: (i) The model is solved numerically. Anderson (2008) documented that the error rate (i.e., absolute errors in solution coefficients divided by their true values) of the Sims’ algorithm is of order 1E-15. Here, the solution algorithm follows Sims except for the handling of indeterminacy. Its accuracy is therefore expected to be comparable. (ii) The integral is replaced by an average. That is, $KL(\theta_0, \theta_1)$ is computed as $(1/(2N)) \sum_{j=-N/2+1}^{N/2} \{\text{tr}(f_{\theta_1}^{-1}(\omega_j) f_{\theta_0}(\omega_j)) - \log \det(f_{\theta_1}^{-1}(\omega_j) f_{\theta_0}(\omega_j)) - n_Y\}$, where N equals 100 and 500 for small and medium scale models. If KL is zero, then the above average is also exactly zero due to the nonnegativity of the summands. Consequently, this introduces approximation errors only when KL is nonzero. (iii) The convergence of the minimization is up to some tolerance level. In light of the above, we tentatively say that the minimized KL is zero if it is below 1E-10. Then, we apply the following three methods to look for evidence contradicting this conclusion.

Method 1. This is due to Qu and Tkachenko (2012, p.116). It computes the maximum absolute and relative deviations of $f_{\theta_1}(\omega)$ from $f_{\theta_0}(\omega)$ across the frequencies: $\max_{\omega \in [0, \pi]} \|f_{\theta_1}(\omega) - f_{\theta_0}(\omega)\|_{\infty}$

and $\{\max_{\omega \in [0, \pi]} \|f_{\theta}(\omega) - f_{\theta_0}(\omega)\|_{\infty}\} / |f_{\theta_0 h}(\omega)|$, where for the latter the denominator is evaluated at the same frequency and element of $f_{\theta_0}(\omega)$ that maximizes the numerator.

Method 2. Compute empirical distance measures using a range of sample sizes. The value should increase consistently with T if $KL(\theta_0, \theta_1)$ is nonzero.

Method 3. Examine whether the data generated by $f_{\theta_0}(\omega)$ are also consistent with $f_{\theta_1}(\omega)$, using

$$X = (2n_Y T)^{-1/2} \sum_{j=1}^{T-1} \left\{ [\log \det(f_{\theta_1}(\omega_j)) + \text{tr}(f_{\theta_1}^{-1}(\omega_j)I(\omega_j))] - [\log \det(f_{\theta_0}(\omega_j)) + n_Y] \right\}, \quad (19)$$

where $I(\omega_j)$ is computed using T observations simulated under $f_{\theta_0}(\omega)$. The summation over the first bracket leads to $-2L_f(\theta_1)$ and the second leads to its mean after imposing $f_{\theta_0}(\omega) = f_{\theta_1}(\omega)$.

Corollary 9 *Let Assumptions 1, 2, 4 and 5 hold. As $T \rightarrow \infty$ with θ_1 and θ_0 staying fixed, we have: (a) if $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$, then $X \xrightarrow{d} N(0, 1)$; (b) if $f_{\theta_1}(\omega) \neq f_{\theta_0}(\omega)$ for some $\omega \in [-\pi, \pi]$, then $T^{-1/2}X \xrightarrow{p} \sqrt{n_Y/2}KL(\theta_0, \theta_1)$, consequently X diverges to $+\infty$.*

In Section 8, we obtain the empirical cdf of X and compare it with $N(0, 1)$, where T is set to a high value of 1E5 for testing power and also to control the effect of the asymptotic approximation.

The three methods are expected to be helpful in confirming observational equivalence when the researcher obtains a near-zero KL value. They all start with the tentative conclusion that the value is in fact zero and then look for evidence against it. The first method can alert us to the situation where there is difference within a narrow band of frequencies which becomes much less visible after taking the average. The second method evaluates the conclusion in light of its implications for empirically relevant sample sizes. It prevents one from making an erroneous conclusion that can interfere with the model's applications in practice. The third method is similar to the second, except that it looks at the entire likelihood distribution and a much greater sample size.

It is feasible to further sharpen the identification results by using higher precision arithmetic. We experiment with a multiprecision computing toolbox for Matlab by Advanpix. The toolbox allows for computing with higher precision (e.g., quadruple precision) at the cost of higher memory usage and longer computational time. It also supports *fminsearch*, a simplex search algorithm for local minimization. Working with this toolbox, we develop the following three-step procedure:

Step 1. The DSGE model solution algorithm is modified so that all the relevant objects are computed as multiprecision entities. In addition, the KL distance is computed using the Gauss-Legendre quadrature with the result saved as a multiprecision entity.

Step 2. Further local minimization is carried out using *fminsearch*. The objective function corresponds to the KL criterion computed in Step 1. The initial values are set to the outputs from *multistart* (i.e., the values reported in the Tables of the paper).

Step 3. The resulting optimizers are compared with their initial values. We check whether the KL values and the parameter vectors are nontrivially different. For the non-identified cases, the KL values are expected to be substantially smaller due to the increased precision. For the identified cases, the KL values should remain essentially the same.

In implementation, the quadruple precision is used and the tolerance level for *fminsearch* is set to 1E-30. Interestingly, to some extent the procedure addresses all three issues raised in the first paragraph of Section 7.2. First, the modification of the model solution algorithm increases the precision from the original 1E-15 to the quadruple precision. Second, by using quadrature rather than a simple average, the KL is also computed more precisely. Finally, the tightening of the tolerance level from 1E-10 to 1E-30 helps the numerical minimizer to move closer to its target value. We thank a referee for pointing towards the high precision arithmetic that has made such progresses possible. The toolbox does not yet allow for global minimization, therefore can not be applied with Matlab’s *ga*, *particleswarm* or *multistart* algorithms. The computational time is also fairly substantial, partly because *fminsearch* utilizes only a single core. For the small scale model considered in Section 8, it takes several hours on a desktop with an Intel 2.4Ghz processor to obtain one result, while for the medium scale model it takes about a day.

7.3 Local identification

The numerical aspects (i) and (ii) remain present. In addition, the derivatives $\partial f_{\theta_0}(\omega_j)/\partial \theta_k$ are computed numerically as $[f_{\theta_0 + \mathbf{e}_k h_k}(\omega_j) - f_{\theta_0}(\omega_j)]/h_k$, where \mathbf{e}_k is a unit vector with the k -th element being 1 and h_k is a step size. Finally, the eigenvalues of the matrix $G(\theta_0)$ are obtained numerically. In the applications, we use the Matlab default tolerance level when determining the rank of $G(\theta_0)$, i.e., $tol = size(G)eps(\|G\|)$, and then further evaluate the conclusion as follows. By Corollary 6 in Qu and Tkachenko (2012), under local identification failure the points on the following curve are observationally equivalent to θ_0 within a local neighborhood of the latter:

$$\partial \theta(v)/\partial v = c(\theta), \theta(0) = \theta_0, \tag{20}$$

where $c(\theta)$ is the eigenvector corresponding to the smallest eigenvalue of $G(\theta)$. We trace out this curve numerically using the Euler method, extending it over a sizable neighborhood, and then apply the three methods in Subsection 7.2 to look for evidence against the identification failure.

Instead of using the one-sided method and the Riemann sum, we can estimate $\partial f_{\theta_0}(\omega_j)/\partial \theta_k$ using the symmetric difference quotient and the integral in (4) using Gaussian quadrature. These alternative numerical methods are applied in Subsection 8.1.1 to examine the result sensitivity.

A referee suggested an intriguing approach of using interval arithmetic to obtain exact bounds on the eigenvalues of $G(\theta_0)$. This is indeed applicable to models whose solutions and derivatives can be obtained analytically. It is an interesting open question whether further developments in the solution methods and in interval arithmetic can make this applicable to general DSGE models.

8 Illustrative applications

This section considers a small and a medium scale DSGE model. The key findings are reported while additional results are available in the supplementary appendix. All the results are based on the full spectrum unless stated otherwise. The empirical distances are all computed at the 5% level.

8.1 An and Schorfheide (2007)

The model is given by: $y_t = E_t y_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}(r_t - E_t \pi_{t+1} - E_t z_{t+1})$, $\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - g_t)$, $r_t = \rho_r r_{t-1} + (1 - \rho_r)\psi_1 \pi_t + (1 - \rho_r)\psi_2(y_t - g_t) + \varepsilon_{rt}$, $g_t = \rho_g g_{t-1} + \varepsilon_{gt}$, $z_t = \rho_z z_{t-1} + \varepsilon_{zt}$, where $\varepsilon_{rt} \sim N(0, \sigma_r^2)$, $\varepsilon_{gt} \sim N(0, \sigma_g^2)$ and $\varepsilon_{zt} \sim N(0, \sigma_z^2)$ are serially and mutually uncorrelated structural shocks. The indeterminacy arises if $\psi_1 < 1 - (1 - \beta)\psi_2/\kappa$. The resulting sunspot shock ϵ_t is one dimensional: $\epsilon_t = M_{r\epsilon}\varepsilon_{rt} + M_{g\epsilon}\varepsilon_{gt} + M_{z\epsilon}\varepsilon_{zt} + \tilde{\epsilon}_t$, where $\tilde{\epsilon}_t \sim N(0, \sigma_\epsilon^2)$. The structural and sunspot parameters are $\theta^D = (\tau, \beta, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z)'$ and $\theta^U = (M_{r\epsilon}, M_{g\epsilon}, M_{z\epsilon}, \sigma_\epsilon)'$.

8.1.1 A contrast between determinacy and indeterminacy

First, consider local identification at the posterior mean of an indeterminacy regime (Column 7 in Table 1). The smallest eigenvalue of $G(\theta_0)$ equals 4.5E-05, well above the Matlab default tolerance level of 6.8E-12. This suggests that θ_0 is locally identified. We further examine this conclusion using the three methods in Section 7. To this end, the curve in (20) is traced out with the step size equal to 1E-04. First, the two deviation measures between $f_\theta(\omega)$ and $f_{\theta_0}(\omega)$ equal 5.5E-04 and 2.4E-04 when $\|\theta - \theta_0\|$ is 0.045, and all reach 1.0E-02 when the norm equals 1.30. The latter values are significantly higher than the numerical errors associated with the Euler method, which should remain near or below 1.0E-04. Second, the empirical distances are computed using points on the curve for $T=80, 150, 200, 1000$. At $\|\theta - \theta_0\| = 0.045$ the resulting values are 0.0524, 0.0531, 0.0535 and 0.0576. They are all above 5% and increase consistently with the sample size. At $\|\theta - \theta_0\| = 1.30$, the values further increase to 0.1723, 0.2316, 0.2702 and 0.6942, respectively. Finally, at the latter θ , the maximum difference between the cdf of X in (19) and that of $N(0, 1)$ equals 0.2433 when $T=1E5$. In summary, all three methods confirm that the model is locally identified at θ_0 .

Next, consider local identification under determinacy (Table 1, Column 10). The smallest eigenvalue of $G(\theta_0^D)$ equals 6.5E-13, below the tolerance level 4.0E-11. This suggests that θ_0^D is locally unidentified. The three verification methods yield the following. First, the deviation measures remain below 1.0E-05 after $\|\theta^D - \theta_0^D\|$ reaches 1.30. Second, the empirical distance at this value remains at 0.0500 when $T=1000$. Third, the maximum difference between the two cdfs equals 0.0144 when $T=1E5$. Thus, they all support the conclusion. This finding is in line with Qu and Tkachenko (2012), who documented the same local identification failure at a different θ_0^D .

We repeat the local identification analysis using the symmetric difference quotient and the Gauss-Legendre quadrature. For the indeterminacy case, the smallest eigenvalue of $G(\theta_0)$, the deviation measures and the empirical distances along the curve are identical to the values reported in the first paragraph up to the decimal places specified there. In the determinacy case, the smallest eigenvalue of $G(\theta_0^D)$ equals 2.6E-14, smaller than the original value. Along the nonidentification curve, the spectral deviation measures remain below 1E-05 (and are slightly smaller than previously) and empirical distance measure equals 0.0500 when $\|\theta - \theta_0\|$ reaches 1.30. Therefore, using the alternative numerical methods leads to the same conclusions.

In the supplementary appendix, we sample from the posterior distributions and check identification at each point. The results show that the above difference is a generic feature of this model. This identification feature has previously been documented in the literature (see Beyer and Farmer, 2004, Lubik and Schorfheide, 2004). In particular, the latter paper illustrated analytically that the sunspot fluctuations generate additional dynamics and therefore contribute to parameter identification. The current paper documents that this can occur in models of empirical relevance.

8.1.2 Global identification under indeterminacy

We minimize $KL(\theta_0, \theta)$ over $\{\theta : \|\theta - \theta_0\|_\infty \geq c\}$ within the parameter bounds specified in the supplementary appendix, with c set to 0.1, 0.5, 1.0 to represent small to fairly large neighborhoods. We note that such c values are not unique, and in practice other values and finer grids can both be considered. The empirical distances are computed for $T = 80, 150, 200, 1000$. The results are reported in Tables 2 and 3, where the bold values highlight the parameters that move the most.

When $c = 0.1$, τ makes the constraint binding. KL equals 5.19E-07. As this is a small value, we compute the empirical distance to gain more insight into the nature of the identification. The resulting values are all above 0.0500 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 difficult to distinguish between them as the empirical distance is only 0.0513 when $T = 150$. When c

is increased to 0.5 and 1.0, τ continues to move the most. The empirical distances remain small, e.g., it is only 0.0672 with $c=1.0$ and $T=150$. In fact, even after τ is set to 10 and $KL(\theta_0, \theta)$ is minimized without further constraints, it only increases to 2.03E-03, with an empirical distance of 0.1960 for $T = 150$. The results thus consistently point to the difficulty in determining τ in the current setting. In addition, the two other verification methods give the following results. Corresponding to the three c values, the two deviation measures equal [0.0156,0.0209], [0.0730,0.1155] and [0.1064,0.2239] and the maximum difference between the cdfs with $T=1E5$ increases from 0.0132 to 0.0143 and 0.0209. In this application, the third verification method is less informative than the first two.

We repeat the above analysis while fixing τ at 2.51. The improvement remains very small: the empirical distance is only 0.0900 with $c=1.0$ and $T=150$. Interestingly, ψ_2 moves the most in all three cases. This suggests that different policy rule parameters can result in near observational equivalence even if they are globally identified. Finally, when both τ and ψ_2 are fixed at their original values, the improvement becomes more noticeable. This process can be continued by fixing additional parameters, to further measure the improvements in the strength of identification. This allows us not only to check global identification and measure the feasibility of distinguishing between models, but also sequentially extract parameters that are responsible for (near) identification failure.

We consider different neighborhood specifications to evaluate the result sensitivity. The results are summarized below while the corresponding parameter values, KL and empirical distances are reported in Tables S1 to S8 in the supplementary appendix. First, we replace $\|\cdot\|_\infty$ with L_1 and L_2 norms. In both cases, the conclusions remain the same, i.e., no observational equivalence is found and the parameters that move the most are still τ, ψ_2, σ_z and $M_{z\epsilon}$. Next, we measure the changes in relative terms, i.e., replacing $\|\theta - \theta_0\|_\infty$ with $\|(\theta - \theta_0)/w(\theta_0)\|_\infty$. Here, $w(\theta_0)$ is chosen to be the lengths of the 90% credible sets reported in Table 1. Beside providing a further verification of the global identifiability, this is also interesting as a reflection on the identification strength in light of the information conveyed by the Bayesian analysis. Again, no observational equivalence is found. The parameters that move the most in relative terms are overall consistent with the above and are, sequentially, β, ψ_2, τ and σ_ϵ . Because the empirical distances are consistently at low values, i.e., the maximum being 0.0825 with $T=150$, the results imply substantially lower distinguishability between parameter values than what one might conclude from reading the posterior credible sets.

In addition, we group the parameters into three subsets and examine them further. The constraint $\{\theta : \|\theta(\text{sel}) - \theta_0(\text{sel})\|_\infty \geq c\}$ is used throughout. First, we subject the monetary policy parameters $(\psi_1, \psi_2, \rho_r, \sigma_r)$ to the constraint while letting the rest vary freely. The empirical distances at $T=150$ for the three c values equal 0.0531, 0.0714, 0.0898. In all cases, ψ_2 makes the

constraint binding. Next, we impose the constraint on parameters related to the exogenous shocks $(\rho_g, \rho_z, \sigma_g, \sigma_z, \sigma_r, M_{r\epsilon}, M_{g\epsilon}, M_{z\epsilon}, \sigma_\epsilon)$. The empirical distances equal 0.0616, 0.0866 and 0.1446. The constraint is binding for σ_z when $c=0.1$, for $M_{z\epsilon}$ when $c=0.5$ and 1.0. Finally, we subject (τ, β, κ) to the constraint. The results are now identical to the first three columns in Tables 2 and 3. Besides making the changes in parameter values more comparable, such an exercise also allows us to assess the identification of a particular set of parameters without making statements about the others.

We present a visualization of $KL(\theta_0, \theta)$ as a function of θ . This can be informative about the challenges and opportunities associated with the global optimization. Because θ consists of 15 parameters, some dimensional reduction is necessary for such visualization to be informative. To this end, the parameters are divided into four subsets that correspond to the four monetary policy parameters, four structural shock parameters, four sunspot parameters, and three behavioral parameters. Within each subset, the two parameters that are relatively weakly identified (chosen in light of the results in Table 2) are varied over a substantial range, while the rest are fixed at some chosen values. This leads to 30 plots, reported in Figures 1 to 4. They illustrate the smoothness, curvature and potential multimodality of the KL surface in the respective lower dimensional spaces specified by the subsets. To ease the visualization, the negative of KL is plotted in all the figures.

Figure 1 illustrates how KL changes with the monetary policy parameters. There, ψ_1 and ψ_2 are varied over $[0.01, 0.9]$ and $[0.01, 5]$, and ρ_r and σ_r are fixed at θ_0 (in Plot (e)) or are perturbed by 0.1 relative to θ_0 . In Plot (e), the changes in KL are quite small over a sizeable neighborhood when ψ_1 and ψ_2 are moved in opposite directions. In other directions, its changes are more visible. When only ρ_r is increased by 0.1 (see Plot (h)) KL remains fairly close to zero, while the remaining seven cases all show clear departures from zero. Overall, the results suggest that KL is a smooth function of the parameters. It can be relatively flat in some directions and have substantial curvatures in others. In addition, all of the nine plots display concavity.

Figures 2 and 3 correspond to the structural shock and sunspot parameters. They are structured in the same way as Figure 1. The nine plots in Figure 2 exhibit concavity and curvatures in all directions. This is consistent with the empirical finding that the structural shock parameters are often fairly adequately estimated. Figure 3 is very different. Although there are still smoothness and visible curvatures, the concavity is lost and clear bimodality is present in six out of the nine subfigures. Such bimodality implies that a standard local optimization algorithm with a single initial value will be inadequate for carrying out the optimization. Figure 4 shows how KL changes with the behavioral parameters. The surfaces are relatively flat in both τ and κ . At the same time, similarly to Figures 1 and 2, they display smoothness and concavity.

In summary, the figures show that there are both challenges (multimodality and the lack of curvature in some dimensions) and opportunities (smoothness and there being substantial curvature in some dimensions) for carrying out the global optimization. In Section 8.1.3, we will provide further details on how the algorithms in Section 7.1 have performed under such features.

8.1.3 Identification of policy rules

This subsection considers the feasibility of distinguishing the original model from two models with the same structure but different policy rules. In the first model, the central bank responds to expected inflation: $r_t = \rho_r r_{t-1} + (1 - \rho_r) \psi_1 E_t \pi_{t+1} + (1 - \rho_r) \psi_2 (y_t - g_t) + \varepsilon_{rt}$. In the second, it reacts to output growth: $r_t = \rho_r r_{t-1} + (1 - \rho_r) \psi_1 \pi_t + (1 - \rho_r) \psi_2 (\Delta y_t + z_t) + \varepsilon_{rt}$. The latter rule is also considered in An and Schorfheide (2007). We denote the alternative models' spectral densities by $h_\phi(\omega)$. The parameters of the original model are fixed at their posterior means, while those of the alternative models are determined by minimizing $KL_{fh}(\theta_0, \phi)$. The results are in Tables 4 and 5.

Consider the expected inflation rule. Under indeterminacy, the minimum of $KL_{fh}(\theta_0, \phi)$ equals 3.38E-14, where ψ_2 changes relative to θ_0 . The small magnitude of KL suggests that the two models are observationally equivalent. The three methods in Section 7 support this conclusion. The empirical distance remains at 0.0500 after T is increased to 1000. The two deviation measures equal 1.32E-06 and 9.56E-07 while the maximum difference between the two cdfs equals 0.0131 when $T=1E5$. Under determinacy, $KL_{fh}(\theta_0^D, \phi^D)$ equals 4.24E-14, where ψ_1 and ψ_2 move relative to θ_0^D . The verification methods yield similar values as above and are omitted. Because the alternative model is unidentified under determinacy, the minimization can yield parameter values that differ from ϕ^D in $(\psi_1, \psi_2, \rho_r, \sigma_r)$. The above observational equivalence is intuitive because the Phillips curve implies a deterministic relationship between $\pi_t, E_t \pi_{t+1}$ and $(y_t - g_t)$.

Consider the output growth rule. Under indeterminacy, $KL_{fh}(\theta_0, \phi)$ is minimized with a value of 1.94E-05. The empirical distances are 0.0583 and 0.0738 when $T=150$ and 1000. The two deviation measures equal 0.0657 and 0.1147, while the maximum difference between the two cdfs equals 0.0151. Under determinacy, $KL_{fh}(\theta_0^D, \phi^D)$ is minimized with a value of 6.00E-05. The empirical distances are 0.0659 and 0.0976 for $T=150$ and 1000. The deviation measures equal 0.0145 and 0.1441, while the difference between the two cdfs equals 0.0205. Therefore, in both cases, we obtain a model that is nearly but not exactly equivalent to the original one.

We examine the local solutions produced by the two-step procedure (*particleswarm* followed by *multistart*) for the 13 cases reported in the columns of Tables 2 and 4. Note that in each case, 110 initial values are used for *multistart*. The first 60 values are taken from the output

of the *particleswarm*, while the remaining 50 values are sampled randomly (therefore are noisy). Throughout, two parameter values are classified as distinct solutions if the sup norm of their difference is no less than 1E-03. Such a classification is consistent with numbers of digits reported in the two tables. The results are as follows. The numbers of distinct local solutions are: 3, 5, 11, 10, 8, 14, 8, 5, 5, 1, 1, 1 and 8. The numbers of initial values that converge to the global optimum are: 12, 99, 85, 65, 100, 94, 55, 50, 52, 110, 108, 110 and 80. Two features emerge. First, the proportions of converged points are high except in the first case. There, 107 initial values converge to a slightly inferior point (with KL being 5.37E-07 as opposed to 5.19E-07), where τ is increased rather than decreased by 0.1 as in Table 2. Second, because no inequality constraints are present for the four cases in Table 4, the proportions of converged points there are even more encouraging. Overall, the results show both the feasibility of the global optimization and also the importance of utilizing multiple initial values.

8.1.4 Results from using high precision arithmetic

This subsection applies the three steps outlined in Section 7.2 to further evaluate the robustness of conclusions on global identification.

First, we consider the minimization of $KL(\theta_0, \theta)$ over $\{\theta : \|\theta - \theta_0\|_\infty \geq c\}$, using the parameters in the nine columns of Table 2 as initial values. The resulting KL and parameter values are all close to those reported in Tables 2 and 3. Specifically, the nine KL values are (compared with the nine columns in Table 3): 5.16E-07, 1.57E-05, 7.54E-05, 1.04E-05, 1.64E-04, 3.34E-04, 5.52E-05, 4.17E-04 and 1.21E-03. Meanwhile, the sup norm differences between the new minimizers and the original θ values are at most 9.82E-04.

Next, we consider the (near) observational equivalence between the policy rules. Using the first two columns of Table 4 as initial values, the minimized KL values are now 1.29E-30 and 4.64E-32, respectively. These are substantially smaller than the original values of 3.38E-14 and 4.24E-14, and give further evidence that the policy rules are indeed observationally equivalent. The sup norm differences between the new minimizers and the original θ values are 4.81E-06 and 4.84E-07. Using the last two columns of Table 4 as initial values, the resulting KL values are 1.94E-05 and 6.00E-05 respectively. This confirms that the policy rules are nearly, but not exactly, observationally equivalent. The sup norm differences between the θ values are 1.78E-04 and 1.33E-06.

In summary, for the two cases where the original procedure detects identification failure (i.e., Panel (a) in Table 4), both KL values are substantially reduced. For the remaining cases, the KL values and the parameter values all remain essentially the same. Therefore, the conclusions remain

the same under the increased precision and the computation of the KL using the quadrature.

8.2 Smets and Wouters (2007)

We carry out the following analysis: (1) study global identification at the posterior mean, (2) detect parameters that are identified weakly, (3) study the feasibility of distinguishing between different policy rules, and (4) compute empirical distances between models with different frictions. The focus is on determinacy as this has been the model’s intended application. The structural parameters are listed in Table 6. The equations and parameter bounds are in the supplementary appendix.

8.2.1 Global identification under determinacy

We minimize KL over $\{\theta : \|\theta - \theta_0\|_\infty \geq c\}$ within the parameter bounds. The results are reported in Panel (a) of Tables 7 and 8. No observational equivalence is found, however, there are models hard to distinguish from the benchmark even with $c=1.0$, where the empirical distance equals 0.1159 when $T=150$. The closest model is always found by moving φ . After φ is fixed at 5.74 (Panel (b)), the improvements for $c=0.1$ and 0.5 are relatively small for typical sample sizes, while substantial changes occur when $c=1.0$. The largest parameter change is always in σ_l . The analysis thus shows that the model is globally identified at the posterior mean after fixing the 5 parameters as in the original study. Also, it shows that among the behavioral parameters, φ and σ_l are identified weakly. The latter is consistent with Smets and Wouters (2007, p.594). In Subsection S.6.1 of the supplementary appendix, verifications using the methods in Section 7 and various sensitivity checks as in Subsection 8.1.2 support the above results.

8.2.2 Identification of policy rules

The original model has the following policy rule $r_t = \rho r_{t-1} + (1 - \rho)(r_\pi \pi_t + r_y (y_t - y_t^*)) + r_{\Delta y}((y_t - y_t^*) - (y_{t-1} - y_{t-1}^*)) + \varepsilon_t^r$. We consider another rule with π_t replaced by $E_t(\pi_{t+1})$. The KL is minimized at $(0.52, 0.80, 0.71, 0.18, 0.58, 5.45, 1.40, 0.70, 1.60, 0.55, 0.64, 0.27, 0.67, 1.35, 2.36, 0.22, 0.08, 0.80, 0.95, 0.22, 0.97, 0.72, 0.17, 0.90, 0.96, 0.45, 0.23, 0.53, 0.45, 0.25, 0.14, 0.24, 0.45, 0.01)'$ with a value of 0.0080. The empirical distance (Table 9) equals 0.4829 when $T=150$. We also consider two alternative rules by letting $r_y=0$ and $r_{\Delta y}=0$, respectively. The minimized KL values equal 0.0499 and 0.1334, and the empirical distances already equal 0.7844 and 0.9965 when $T=80$. Therefore, unlike in Subsection 8.1.3, here no near observational equivalence is found. This result illustrates how identification of policy rules changes in important ways when moving from small to medium scale models. Further checks in the supplementary appendix support the above conclusion, see Subsection S.6.2.

8.2.3 Identification of frictions

This subsection examines how the dynamics of the model are affected when the frictions are reduced substantially from their original values. Here the relevant frictions 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 them one at a time to dampen the effect of the corresponding friction, and then search for the closest model in terms of the KL criterion. Table 10 contains the results.

Consider the nominal frictions. Reducing ξ_p and ξ_w to 0.10 both lead to an empirical distance of 1.00 when $T=150$. This suggests that these two frictions are very important for generating the dynamic properties. A lower ξ_p leads to a large increase in ι_p from 0.24 to 0.99. The price mark-up shock process becomes amplified. A lower ξ_w leads to an increase in ι_w from 0.58 to 0.99. The wage mark-up process becomes amplified. Interestingly, φ and σ_l decrease sharply in both cases. The remaining two frictions, ι_p and ι_w , are also relevant but less important. Reducing ι_p to 0.01 yields an empirical distance of 0.46 when $T=150$. Its KL value is also the lowest among all the cases. Leaving out either friction does not have any noticeable impact on the other parameters.

The results regarding the four real frictions are as follows. (1) Reducing φ from 5.74 to 1.00 leads to an empirical distance of 1.00 for $T=150$. The persistence and volatility of the investment shock process increase strongly. (2) Reducing λ from 0.71 to 0.10 also yields an empirical distance of 1.00 for $T=150$. Here, ψ increases, while φ and 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% leads to an empirical distance of 1.00 for $T=150$. The standard deviation of the technology shock increases and the latter also becomes more persistent. In summary, all four real frictions are important in generating the dynamic properties of the model with ψ being the least so. The conclusions reached regarding the eight frictions are consistent with the findings in Smets and Wouters (2007, Table 4).

Remark 4 *We conduct two further checks. (The details are reported in Subsections S4.2, S4.3, S6.3 and S6.4 of the supplementary appendix.) First, to examine whether the empirical distance measure closely tracks the exact power of the time domain likelihood ratio test, we compute the latter using simulations at the parameter values reported in Tables 2 and 7 as examples for the small and medium scale model respectively. The resulting values are quite close to those in Tables 3 and*

8. The maximum difference across all the cases is only 0.0037. Second, to examine the extent of asymmetry present, we compute the empirical distances using the values in Tables 2,4,7,10 and Subsection 8.2.2 with the benchmark and alternative models switched. The resulting values are fairly close to those in Tables 3,5,8,10 and 9. The maximum difference equals 0.0614 concerning Tables 9 and 10 and 0.0056 regarding the rest.

8.2.4 Results from using high precision arithmetic

This subsection carries out the same analysis as in Subsection 8.1.4. First, the minimization of $KL(\theta_0, \theta)$ over $\{\theta : \|\theta - \theta_0\|_\infty \geq c\}$ leads to essentially the same KL and parameter values as those in Tables 7 and 8. Specifically, the six KL values are (compared with the six columns in Table 8): 8.15E-06, 1.86E-04, 6.66E-04, 2.87E-05, 5.86E-04 and 1.88E-03. Meanwhile, the differences between the new minimizers and the original θ values are at most 4.66E-04. Second, for the monetary policy rules, the KL values are 0.0080, 0.0499 and 0.1334. The differences between the new minimizers and the original θ values are at most 5.28E-03. Finally, for the frictions, the KL values are (compared with the eight columns in Table 10): 0.3015, 0.2433, 0.0078, 0.0284, 0.1133, 0.1712, 0.0278 and 0.0923. The differences between the parameter values are at most 3.04E-02. Overall, when KL is higher, the difference between approximating it using a simple average or the quadrature also tends to be higher. This is why the differences in the parameter values that minimize the KL are higher in the final case relative to the previous cases. In summary, for all cases considered, the conclusions remain qualitatively the same under the increased precision and the computation of the KL using the quadrature.

9 Conclusion

This paper developed necessary and sufficient conditions for global identification in DSGE models that encompass both determinacy and indeterminacy. It also proposed a measure for the empirical closeness between DSGE models. The latter enables us to interpret identification not merely as a zero-one phenomenon, but also relative to the samples sizes encountered in empirical applications.

Bayesian analysis of DSGE models has rapidly progressed, evidenced both by the increased theoretical treatments (An and Schorfheide, 2007, Fernández-Villaverde, 2010, Herbst and Schorfheide, 2015) and by the number of applications. The progresses from the frequentist perspective have remained much limited (surveyed in Fernández-Villaverde, Rubio-Ramírez and Schorfheide, 2015). If a researcher wishes to carry out global identification analysis and model comparisons while (temporarily) setting aside the prior, few if any methods are available. We hope this paper can be one

step in such a direction. The applications of the methods developed here, in particular to the identification contrast between determinacy and indeterminacy, the (near) observational equivalence of different monetary policy rules in small scale models, and the assessment of various frictions in medium scale models, suggest that such endeavors are feasible and can be potentially fruitful.

References

- ADVANPIX LLC. (2016): “Multiprecision Computing Toolbox for MATLAB,” Yokohama, Japan.
- 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.
- ANDERSON, G. S. (2008): “Solving Linear Rational Expectations Models: A Horse Race,” *Computational Economics*, 31, 95–113.
- 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 (2014): “Choosing the Variables to Estimate Singular DSGE Models,” *Journal of Applied Econometrics*, 29, 1099–1117.

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.

DAVIES, R. B. (1973): “Asymptotic Inference in Stationary Gaussian Time-Series,” *Advances in Applied Probability*, 5, 469–497.

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.

DZHAPARIDZE, K. (1986): *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer-Verlag.

FERNÁNDEZ-VILLAYERDE, J. (2010): “The Econometrics of DSGE Models,” *Journal of the Spanish Economic Association*, 1, 3–49.

FERNÁNDEZ-VILLAVERDE, J., AND J. F. RUBIO-RAMÍREZ (2004): “Comparing Dynamic Equilibrium Models to Data: a Bayesian Approach,” *Journal of Econometrics*, 123, 153–187.

FERNÁNDEZ-VILLAVERDE, J., J. F. RUBIO-RAMÍREZ, AND F. SCHORFHEIDE (2015): “Solution and Estimation Methods for DSGE Models,” *Handbook of Macroeconomics, Volume 2 (forthcoming)*.

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.

HERBST, E. P., AND F. SCHORFHEIDE (2015): *Bayesian Estimation of DSGE Models*. Princeton University Press.

HOSOYA, Y., AND M. TANIGUCHI (1982): “A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes,” *The Annals of Statistics*, 10, 132–153.

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

KOCIECKI, A., AND M. KOLASA (2013): “Global Identification of Linearized DSGE Models,” *Working Paper 170, National Bank of Poland*.

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.

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. Springer, 3 edn.

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. (2014): “Inference in Dynamic Stochastic General Equilibrium Models with Possible Weak Identification,” *Quantitative Economics*, 5, 457–494.

QU, Z., AND D. TKACHENKO (2012): “Identification and Frequency Domain Quasi-maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models,” *Quantitative Economics*, 3, 95–132.

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

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.

Applying the QZ decomposition to (1), we have $Q^*\Lambda Z^* = \Gamma_0$, $Q^*\Omega Z^* = \Gamma_1$, where Q and Z are unitary, Λ and Ω are upper triangular. Let $w_t = Z^*S_t$ and premultiply (1) by Q :

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} (\Psi\varepsilon_t + \Pi\eta_t), \quad (\text{A.1})$$

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. 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

$$Q_2.\Pi\eta_t = -Q_2.\Psi\varepsilon_t \text{ for all } t > 0. \quad (\text{A.2})$$

The condition (A.2) determines $Q_2.\Pi\eta_t$ as a function of ε_t . However, it may be insufficient to determine $Q_1.\Pi\eta_t$, in which case it will lead to indeterminacy.

Because the rows of $Q_2.\Pi$ can be linearly dependent, Sims (2002) and Lubik and Schorfheide (2003) suggested to work with its SVD to isolate the effective restrictions imposed on η_t :

$$Q_2.\Pi = [U_{.1} \ U_{.2}] \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{.1}^* \\ V_{.2}^* \end{bmatrix} = U_{.1}D_{11}V_{.1}^*, \quad (\text{A.3})$$

where $[U_{.1} \ U_{.2}]$ and $[V_{.1} \ V_{.2}]$ are unitary matrices and D_{11} is nonsingular. The submatrices $U_{.1}$ and $V_{.1}^*$ are unique up to multiplication by a unit-phase factor $\exp(i\varphi)$ (for the real case, up to sign). The spaces spanned by $U_{.2}$ and $V_{.2}$ 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 $V_{.2}$ when implementing the relevant procedures. Note that matrices with the same column space have the same unique reduced column echelon form.

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

$$U_{.1}D_{11}V_{.1}^*\eta_t = -Q_2.\Psi\varepsilon_t \text{ for all } t > 0. \quad (\text{A.4})$$

Premultiplying (A.4) by the conjugate transpose of $[U_{.1} \ U_{.2}]$ does not alter the restrictions because the latter is nonsingular. Thus, (A.4) is equivalent to (using $U_{.1}^*U_{.1} = I$ and $U_{.2}^*U_{.1} = 0$) the following two sets of equations: $D_{11}V_{.1}^*\eta_t = -U_{.1}^*Q_2.\Psi\varepsilon_t$ and $0 = -U_{.2}^*Q_2.\Psi\varepsilon_t$ for all $t > 0$. The second set of equations places no restrictions on η_t . The first set 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

$$\{p + v : v \text{ is any solution to } Ax = 0 \text{ and } p \text{ is a specific solution to } Ax = b\}. \quad (\text{A.5})$$

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 $\eta_t = V_{.2}\varepsilon_t$ with ε_t being an arbitrary vector conformable with $V_{.2}$. Therefore, the full set of solutions to (A.4) can be represented as

$$\{\eta_t : \eta_t = -V_{.1}D_{11}^{-1}U_{.1}^*Q_2.\Psi\varepsilon_t + V_{.2}\varepsilon_t \text{ with } E_{t-1}\varepsilon_t = 0\}. \quad (\text{A.6})$$

The restriction $E_{t-1}\varepsilon_t = 0$ 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).

We now provide some computational details on how to use (A.6) to solve for S_t in (1) as in Sims (2002). Define Φ as the projection 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 $Q_1.\Pi - \Phi Q_2.\Pi = Q_1.\Pi - Q_1.\Pi V_{.1}V_{.1}^* = Q_1.\Pi(I - V_{.1}V_{.1}^*)$, which equals zero under determinacy. Multiplying (A.1) by

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

and imposing the restrictions (A.2):

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi\Lambda_{22} \\ 0 & I \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_1. - \Phi Q_2. \\ 0 \end{bmatrix} (\Psi\varepsilon_t + \Pi\eta_t).$$

Further, using the expression (A.6),

$$\begin{aligned} (Q_1. - \Phi Q_2.)\Pi\eta_t &= (Q_1.\Pi - \Phi Q_2.\Pi)(-V_{.1}D_{11}^{-1}U_{.1}^*Q_2.\Psi\varepsilon_t + V_{.2}\varepsilon_t) \\ &= -Q_1.\Pi(I - V_{.1}V_{.1}^*)V_{.1}D_{11}^{-1}U_{.1}^*Q_2.\Psi\varepsilon_t + Q_1.\Pi(I - V_{.1}V_{.1}^*)V_{.2}\varepsilon_t. \end{aligned}$$

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

$$\begin{aligned} \begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi\Lambda_{22} \\ 0 & I \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} &= \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_1. - \Phi Q_2. \\ 0 \end{bmatrix} \Psi\varepsilon_t \\ &\quad + \begin{bmatrix} Q_1.\Pi(I - V_{.1}V_{.1}^*) \\ 0 \end{bmatrix} V_{.2}\varepsilon_t. \end{aligned}$$

Call the most left hand side matrix G_0 . Multiply the above equation by ZG_0^{-1} and using $w_t = Z^*S_t$, we obtain $S_t = \Theta_1 S_{t-1} + \Theta_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$, where

$$\begin{aligned} \Theta_1 &= ZG_0^{-1} \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & 0 \end{bmatrix} Z^*, \quad \Theta_\varepsilon = ZG_0^{-1} \begin{bmatrix} Q_1. - \Phi Q_2. \\ 0 \end{bmatrix} \Psi, \\ \Theta_\epsilon &= ZG_0^{-1} \begin{bmatrix} Q_1.\Pi(I - V_{.1}V_{.1}^*) \\ 0 \end{bmatrix} V_{.2}. \end{aligned}$$

Further, applying the triangular structure of G_0^{-1} , the above matrices can be represented as $\Theta_1 = Z_{.1}\Lambda_{11}^{-1}[\Omega_{11} \quad \Omega_{12} - \Phi\Omega_{22}]Z^*$, $\Theta_\varepsilon = Z_{.1}\Lambda_{11}^{-1}(Q_1. - \Phi Q_2.)\Psi$ and $\Theta_\epsilon = Z_{.1}\Lambda_{11}^{-1}Q_1.\Pi(I - V_{.1}V_{.1}^*)V_{.2}$, where $Z_{.1}$ includes the first block of columns of Z conformable with Λ_{11} .

Appendix B. Proofs

For notation, let $|A|$ and $\|A\|$ be the Euclidean and vector induced norms of an n dimensional square matrix A . Let $\text{tr}(A)$ be its trace and $\text{vec}(A)$ be its vectorization obtained by stacking the columns on top of one another. The following relationships are useful for the subsequent analysis (see, e.g., p. 73 in Dzhaparidze (1986)):

$$|A| \leq n^{1/2} \|A\|, \quad |\text{tr}(AB)| \leq |A| |B| \quad \text{and} \quad |AB| \leq \|A\| |B|. \quad (\text{B.1})$$

Proof of Theorem 1. See the supplementary appendix.

Proof of Theorem 2. For any $\omega \in [-\pi, \pi]$, $f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)$ has an eigenvalue decomposition with all the eigenvalues being 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: $f_{\theta_1}(\omega) = A^*(\omega)A(\omega)$ and $f_{\theta_0}(\omega) = B^*(\omega)B(\omega)$. 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 $A(\omega)^{-1}$, we obtain

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

The right hand side is a positive definite matrix. It has a well defined eigen decomposition with strictly positive eigenvalues. Denote this eigen decomposition by $\tilde{V}^{-1}(\omega)\Lambda(\omega)\tilde{V}(\omega)$. Then

$$\begin{aligned} f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega) &= A(\omega)^{-1}\tilde{V}^{-1}(\omega)\Lambda(\omega)\tilde{V}(\omega)A(\omega) \\ &= V^{-1}(\omega)\Lambda(\omega)V(\omega) \end{aligned} \quad (\text{B.2})$$

with $V(\omega) = \tilde{V}(\omega)A(\omega)$. It follows that $f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega)$ has an eigenvalue decomposition with the eigenvalues given by the diagonal elements of $\Lambda(\omega)$.

Let $\lambda_j(\omega)$ denote the j -th largest eigenvalue. Then,

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

For any $\lambda_j(\omega) > 0$, the following inequality always holds:

$$\lambda_j(\omega) - \log \lambda_j(\omega) - 1 \geq 0, \quad (\text{B.3})$$

where the equality obtains if and only if $\lambda_j(\omega) = 1$. Integrating the inequality over $[-\pi, \pi]$, we obtain $\text{KL}(\theta_0, \theta_1) \geq 0$, where the equality holds if and only if $\Lambda(\omega) = I$ for all $\omega \in [-\pi, \pi]$. By (B.2), $\Lambda(\omega) = I$ is equivalent to $f_{\theta_1}^{-1}(\omega)f_{\theta_0}(\omega) = V^{-1}(\omega)V(\omega) = I$ 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 Corollary 1. By (B.3) and the nonnegativity of $W(\omega)$, the following inequality always holds: $W(\omega)(\lambda_j(\omega) - \log \lambda_j(\omega) - 1) \geq 0$, where the equality holds if and only if $\lambda_j(\omega) = 1$ or $W(\omega) = 0$. Integrating this inequality over $[-\pi, \pi]$, we obtain $KL^W(\theta_0, \theta_1) \geq 0$, where the equality holds if and only if $\Lambda(\omega) = I$ for all the ω with $W(\omega) > 0$. This is equivalent to $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all the ω with $W(\omega) > 0$.

Proof of Corollary 2. The result follows from Theorem 2 and the positive definiteness of $f_{\theta_1}^{-1}(0)$.

Proof of Corollary 3. The proof is identical to that of Theorem 2 with $h_\phi(\omega)$ replacing $f_{\theta_1}(\omega)$.

Proof of Theorem 3. We establish the limiting distribution under the null hypothesis that $f_{\theta_0}(\omega)$ is the true spectrum. The proof under the alternative hypothesis is essentially the same.

First, by Assumption 4 and the Lipschitz condition in Assumption 5, we have

$$\begin{aligned} & T^{1/2} \left(\frac{1}{2T} \sum_{j=1}^{T-1} \{ \log \det(h_{\phi_0}^{-1}(\omega_j) f_{\theta_0}(\omega_j)) - \text{tr}(h_{\phi_0}^{-1}(\omega_j) f_{\theta_0}(\omega_j)) + n_Y \} + KL_{fh}(\theta_0, \phi_0) \right) \\ &= O(T^{-\beta+1/2}) = o(1) \quad \text{because } \beta > 1/2. \end{aligned} \quad (\text{B.4})$$

Second, we have

$$\begin{aligned} & \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \text{tr} \left\{ \left(f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right) (I(\omega_j) - f_{\theta_0}(\omega_j)) \right\} \\ &= \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \text{vec} \left\{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right\}^* \text{vec} \{ I(\omega_j) - f_{\theta_0}(\omega_j) \}. \end{aligned} \quad (\text{B.5})$$

By Lemma A.3.3(2) of Hosoya and Taniguchi (1982, p.150), (B.5) satisfies a multivariate central limit theorem. Its limiting variance is given by

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{2T} \sum_{j=1}^{T-1} \text{vec} \left\{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right\}^* (f_{\theta_0}(\omega_j)' \otimes f_{\theta_0}(\omega_j)) \text{vec} \left\{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right\} \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \right\} d\omega. \end{aligned} \quad (\text{B.6})$$

The result follows from combining (B.4), (B.5) and (B.6).

Proof of Corollary 4. The proof proceeds by applying Taylor expansions to analyze $KL_{fh}(\theta_0, \phi_0)$, $KL_{hf}(\phi_0, \theta_0)$, $V_{fh}(\theta_0, \phi_0)$ and $V_{hf}(\phi_0, \theta_0)$ in the definition of $p_{fh}(\theta_0, \phi_0, \alpha, T)$ in (12)-(13). Because all the quantities are evaluated with $f = h$, we omit the subscripts in the expressions.

Apply a second order expansion around θ_0 :

$$KL(\theta_0, \theta_T) = \frac{\partial KL(\theta_0, \theta_T)}{\partial \theta'} \Big|_{\theta_0} (\theta_T - \theta_0) + \frac{1}{2} (\theta_T - \theta_0)' \frac{\partial^2 KL(\theta_0, \theta_T)}{\partial \theta \partial \theta'} \Big|_{\theta_0} (\theta_T - \theta_0) + o(T^{-1}). \quad (\text{B.7})$$

Because

$$\partial KL(\theta_0, \theta_T) / \partial \theta = (4\pi)^{-1} \int_{-\pi}^{\pi} (\partial \text{vec} f_{\theta_0}(\omega) / \partial \theta')^* \{ f_{\theta_T}^{-1}(\omega)' \otimes f_{\theta_T}^{-1}(\omega) \} \text{vec}(f_{\theta_T}(\omega) - f_{\theta_0}(\omega)) d\omega,$$

the first term in the expansion (B.7) is in fact identically zero. Differentiating the right hand side and evaluating it at $\theta_T = \theta_0$, we obtain: $\partial^2 KL(\theta_0, \theta_T) / \partial \theta \partial \theta' \Big|_{\theta_0} = I(\theta_0)$. The third order derivative of $KL(\theta_0, \theta)$ with respect to the j, k and l -th elements of θ can be written as

$$(4\pi)^{-1} \int_{-\pi}^{\pi} \text{tr} (D_{jkl}(\omega)) d\omega, \quad (\text{B.8})$$

where $D_{jkl}(\omega)$ is a summation of a finite number of terms, each being the product of some of the following components:

$$f_{\theta_0}(\omega), f_{\theta}^{-1}(\omega), \frac{\partial f_{\theta}(\omega)}{\partial \theta_j}, \frac{\partial f_{\theta}(\omega)}{\partial \theta_k}, \frac{\partial f_{\theta}(\omega)}{\partial \theta_l}, \frac{\partial^2 f_{\theta}(\omega)}{\partial \theta_j \partial \theta_k}, \frac{\partial^2 f_{\theta}(\omega)}{\partial \theta_j \partial \theta_l}, \frac{\partial^2 f_{\theta}(\omega)}{\partial \theta_k \partial \theta_l}, \frac{\partial^3 f_{\theta}(\omega)}{\partial \theta_j \partial \theta_k \partial \theta_l}.$$

These components are all finite because of Assumptions 2(ii), 4, and that $f_{\theta}(\omega)$ has finite first to third order derivatives in an open neighborhood of θ_0 . Therefore, (B.8) is finite. This verifies that the remainder term in (B.7) is $o(1)$. Consequently,

$$KL(\theta_0, \theta_T) = (1/2) (\theta_T - \theta_0)' I(\theta_0) (\theta_T - \theta_0) + o(T^{-1}). \quad (\text{B.9})$$

Similarly, we can show that $KL(\theta_T, \theta_0)$ also equals the right hand side of (B.9).

Now consider $V(\theta_0, \theta_T)$. Its first order derivative equals

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \text{vec} f_{\theta_T}(\omega)}{\partial \theta'} \right)^* \left\{ f_{\theta_T}^{-1}(\omega)' \otimes f_{\theta_T}^{-1}(\omega) \right\} \text{vec} \left\{ \left[I - f_{\theta_0}(\omega) f_{\theta_T}^{-1}(\omega) \right] f_{\theta_0}(\omega) \right\} d\omega,$$

which equals zero when evaluated at $\theta_T = \theta_0$. Differentiating the preceding display and evaluating it at $\theta_T = \theta_0$, we obtain: $\partial^2 V(\theta_0, \theta_T) / \partial \theta \partial \theta' |_{\theta_0} = 2I(\theta_0)$. Its third order derivative can be analyzed in the same way as that of (B.7) and is $o(1)$. Therefore,

$$V(\theta_0, \theta_T) = (\theta_T - \theta_0)' I(\theta_0) (\theta_T - \theta_0) + o(T^{-1}).$$

Similarly, we can show that $V(\theta_T, \theta_0)$ also equals the right hand side of the preceding display.

Applying these results to (12)-(13):

$$\begin{aligned} & p_{ff}(\theta_0, \theta_T, \alpha, T) \\ = & \Pr \left(Z > \frac{\left[-T^{-1/2} \delta' I(\theta_0) \delta / 2 + T^{-1/2} \sqrt{\delta' I(\theta_0) \delta} z_{1-\alpha} \right] - T^{-1/2} \delta' I(\theta_0) \delta / 2 + o(T^{-1/2})}{T^{-1/2} \sqrt{\delta' I(\theta_0) \delta} + o(T^{-1/2})} \right). \end{aligned}$$

Rearranging the terms on the right hand side leads to the desired result.

Proof of Corollary 5. The proof consists of showing that the leading terms in second order expansions of the time and frequency domain likelihood ratios are asymptotically equivalent. Because these terms determine the respective asymptotic local power functions, the asymptotic powers of these two tests must also coincide. The result then follows because $\Pr(Z > z_{1-\alpha} - \sqrt{\delta' I(\theta_0) \delta})$ is the asymptotic local power function of the frequency domain test against θ_T .

The time domain likelihood ratio satisfies

$$\begin{aligned} \mathcal{L}(\theta_T) - \mathcal{L}(\theta_0) &= T^{-1/2} \frac{\partial \mathcal{L}(\theta_0)}{\partial \theta'} \delta + \frac{1}{2} \delta' \left(T^{-1} \frac{\partial^2 \mathcal{L}(\theta_0)}{\partial \theta \partial \theta'} \right) \delta + o_p(1) \\ &= T^{-1/2} \frac{\partial \mathcal{L}(\theta_0)}{\partial \theta'} \delta + \frac{1}{2} \delta' \left\{ E \left(T^{-1} \frac{\partial^2 \mathcal{L}(\theta_0)}{\partial \theta \partial \theta'} \right) \right\} \delta + o_p(1). \end{aligned} \quad (\text{B.10})$$

The remainder term in the first line is $o_p(1)$ because the third order derivative of $T^{-1} \mathcal{L}(\theta)$ is uniformly bounded over an open neighborhood of θ_0 . The latter is proved in Lemma S.1 (iv). The

expectation in (B.10) arises because the average Hessian converges in probability to its expectation. This holds because, for all $1 \leq j, k \leq p + q$,

$$\begin{aligned} \text{Var} \left(T^{-1} \frac{\partial^2 \mathcal{L}(\theta_0)}{\partial \theta_j \partial \theta_k} \right) &= \frac{1}{4T^2} \text{Var} \left(Y' \frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} Y \right) = \frac{1}{2T^2} \text{tr} \left(\frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} \Omega_{\theta_0} \frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} \Omega_{\theta_0} \right) \quad (\text{B.11}) \\ &\leq \frac{n_Y}{2T} \left\| \frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} \right\|^2 \|\Omega_{\theta_0}\|^2 = O(T^{-1}). \end{aligned}$$

In (B.11), θ equals either θ_0 (under the null hypothesis) or θ_T (under the alternative hypothesis), the second equality follows from properties of quadratic forms, and the inequality follows because of (B.1). We next study the first and second order terms in the expansion (B.10).

The j -th element of $T^{-1/2} \partial \mathcal{L}(\theta_0) / \partial \theta'$ equals

$$T^{-1/2} \frac{\partial \mathcal{L}(\theta_0)}{\partial \theta_j} = \frac{1}{2T^{1/2}} \frac{\partial \log \det \Omega_{\theta_0}^{-1}}{\partial \theta_j} - \frac{1}{2T^{1/2}} Y' \frac{\partial \Omega_{\theta_0}^{-1}}{\partial \theta_j} Y \equiv (\mathcal{A}) + (\mathcal{B}).$$

The frequency domain counterpart is given by

$$T^{-1/2} \frac{\partial L_f(\theta_0)}{\partial \theta_j} = \frac{1}{2T^{1/2}} \sum_{i=1}^{T-1} \frac{\partial \log \det f_{\theta_0}^{-1}(\omega_i)}{\partial \theta_j} - \frac{1}{2T^{1/2}} \sum_{i=1}^{T-1} \text{tr} \left\{ I(\omega_i) \frac{\partial f_{\theta_0}^{-1}(\omega_i)}{\partial \theta_j} \right\} \equiv (\mathcal{A}) + (\mathcal{B}). \quad (\text{B.12})$$

By Lemma S.1 (i)-(ii), $(\mathcal{A}) = (\mathcal{A}) + o(1)$ and $(\mathcal{B}) = (\mathcal{B}) + o_p(1)$. Therefore, these two first order derivatives are asymptotically equivalent.

The (j, k) -th element of $E(T^{-1} \partial^2 \mathcal{L}(\theta_0) / \partial \theta \partial \theta')$ is given by

$$\begin{aligned} E \left(T^{-1} \frac{\partial^2 \mathcal{L}(\theta_0)}{\partial \theta_j \partial \theta_k} \right) &= \frac{1}{2T} \frac{\partial^2 \log \det \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} - \frac{1}{2T} E \left(Y' \frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} Y \right) \\ &= \frac{1}{2T} \text{tr} \left(\frac{\partial \Omega_{\theta_0}}{\partial \theta_j} \frac{\partial \Omega_{\theta_0}^{-1}}{\partial \theta_k} \right) + \frac{1}{2T} \text{tr} \left((\Omega_{\theta_0} - \Omega_{\theta_T}) \frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} \right) \\ &= \frac{1}{2T} \text{tr} \left(\frac{\partial \Omega_{\theta_0}}{\partial \theta_j} \frac{\partial \Omega_{\theta_0}^{-1}}{\partial \theta_k} \right) + o(1), \quad (\text{B.13}) \end{aligned}$$

where the second equality follows because of the chain rule, θ equals either θ_0 or θ_T , and the last equality holds because

$$T^{-1} \left| \text{tr} \left((\Omega_{\theta_0} - \Omega_{\theta_T}) \frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} \right) \right| \leq n_Y \|(\Omega_{\theta_0} - \Omega_{\theta_T})\| \left\| \frac{\partial^2 \Omega_{\theta_0}^{-1}}{\partial \theta_j \partial \theta_k} \right\| = O(T^{-1/2}).$$

Meanwhile, the frequency domain counterpart is given by

$$E \left(T^{-1} \frac{\partial^2 L_f(\theta_0)}{\partial \theta_j \partial \theta_k} \right) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left(f_{\theta_0}^{-1}(\omega) \frac{\partial f_{\theta_0}(\omega)}{\partial \theta_j} f_{\theta_0}^{-1}(\omega) \frac{\partial f_{\theta_0}(\omega)}{\partial \theta_k} \right) d\omega + o(1). \quad (\text{B.14})$$

By Lemma S.1 (iii), the leading term of (B.14) is asymptotically equivalent to that of (B.13). Therefore, the two second order terms are also asymptotically equivalent.

The frequency domain likelihood ratio can be represented as

$$\frac{\delta'}{2T^{1/2}} \sum_{j=1}^{T-1} \left(\frac{\partial \text{vec } f_{\theta_0}(\omega_j)}{\partial \theta'} \right)^* \{f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j)\} \text{vec}(I(\omega_j) - f_{\theta_0}(\omega_j)) - \frac{1}{2} \delta' I(\theta_0) \delta + o_p(1). \quad (\text{B.15})$$

The first two terms are from combining (B.12) and (B.14). The remainder term is $o_p(1)$ because $T^{-1} \partial^3 L_f(\theta) / \partial \theta_j \partial \theta_k \partial \theta_l = O_p(1)$ holds uniformly over an open neighborhood of θ_0 . Under the null hypothesis of $\theta = \theta_0$, the first term in (B.15) is asymptotically distributed as $N(0, \delta' I(\theta_0) \delta)$. Under the alternative hypothesis of $\theta = \theta_T$, it is asymptotically distributed as $N(\delta' I(\theta_0) \delta, \delta' I(\theta_0) \delta)$. Consequently, (B.15) is asymptotically distributed as $N(-\frac{1}{2} \delta' I(\theta_0) \delta, \delta' I(\theta_0) \delta)$ and $N(\frac{1}{2} \delta' I(\theta_0) \delta, \delta' I(\theta_0) \delta)$ under the two hypotheses, respectively. From these two distributions, it follows that the asymptotic power function is given by $\Pr(Z > z_{1-\alpha} - \sqrt{\delta' I(\theta_0) \delta})$.

Proof of Corollary 6. The main technical challenge is that the departure from the null model, i.e., $\text{vec}(h_{\phi_T}(\omega) - f_{\theta_0}(\omega))$, lies generally outside of the column space of $\partial \text{vec } f_{\theta}(\omega) / \partial \theta'$. Consequently, the Taylor expansion argument in (B.10) is no longer applicable. Instead, we need to directly work with quantities such as $f_{\theta_0}^{-1}(\omega) - h_{\phi_T}^{-1}(\omega)$ and $\Omega_f^{-1} - \Omega_h^{-1}$. The former is relatively straightforward because the matrices are of finite dimensions as $T \rightarrow \infty$. The latter is more involved as the matrix dimensions diverge to infinity. The developments in Dzhaparidze (1986, Chapter 1), who showed the asymptotic equivalence between the time and frequency domain likelihood ratios in the context of univariate time series models, turn out to be quite useful for the current analysis.

Consider the first result of the Corollary. We have

$$\begin{aligned} h_{\phi_T}^{-1}(\omega) f_{\theta_0}(\omega) &= \left[f_{\theta_0}(\omega) + T^{-1/2} \delta(\omega) \right]^{-1} f_{\theta_0}(\omega) \\ &= I - T^{-1/2} f_{\theta_0}^{-1}(\omega) \delta(\omega) + T^{-1} [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 + o(T^{-1}) \end{aligned} \quad (\text{B.16})$$

and, by Lemma A1.1 in Dzhaparidze (1986, p.74),

$$\begin{aligned} \log \det(h_{\phi_T}^{-1}(\omega) f_{\theta_0}(\omega)) &= -\log \det \left[I + T^{-1/2} f_{\theta_0}^{-1}(\omega) \delta(\omega) \right] \\ &= -\text{tr} \left\{ T^{-1/2} f_{\theta_0}^{-1}(\omega) \delta(\omega) - (2T)^{-1} [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} + o(T^{-1}), \end{aligned} \quad (\text{B.17})$$

where the $o(T^{-1})$ terms hold uniformly over $\omega \in [-\pi, \pi]$. Apply these two results:

$$\begin{aligned} KL_{fh}(\theta_0, \phi_T) &= (8\pi T)^{-1} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega + o(T^{-1}), \\ V_{fh}(\theta_0, \phi_T) &= (4\pi T)^{-1} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega + o(T^{-1}). \end{aligned} \quad (\text{B.18})$$

Similarly, we can show that $KL_{hf}(\phi_T, \theta_0) = (8\pi T)^{-1} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega + o(T^{-1})$ and $V_{hf}(\phi_T, \theta_0) = (4\pi T)^{-1} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega + o(T^{-1})$. The first result of the Corollary follows by applying these four expressions to (12)-(13).

Consider the second result in the Corollary. We structure the proof as in Corollary 5. That is, we show that the time and frequency domain likelihood ratios are asymptotically equivalent and that the power function of the frequency domain likelihood ratio is given by the stated formula.

In the frequency domain:

$$L_h(\phi_T) - L_f(\theta_0) = -\frac{1}{2} \sum_{j=1}^{T-1} \log \det(f_{\theta_0}^{-1}(\omega_j) h_{\phi_T}(\omega_j)) + \frac{1}{2} \sum_{j=1}^{T-1} \text{tr} \left\{ (f_{\theta_0}^{-1}(\omega_j) - h_{\phi_T}^{-1}(\omega_j)) I(\omega_j) \right\}.$$

The first term on the right hand side is already analyzed in (B.17), while the second term equals

$$\begin{aligned} & \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \text{tr} \left\{ f_{\theta_0}^{-1}(\omega_j) \delta(\omega_j) f_{\theta_0}^{-1}(\omega_j) I(\omega_j) \right\} + \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \text{tr} \left\{ f_{\theta_0}^{-1}(\omega_j) \delta(\omega_j) (h_{\phi_T}^{-1}(\omega_j) - f_{\theta_0}^{-1}(\omega_j)) I(\omega_j) \right\} \\ = & \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \text{tr} \left\{ f_{\theta_0}^{-1}(\omega_j) \delta(\omega_j) f_{\theta_0}^{-1}(\omega_j) I(\omega_j) \right\} - \frac{1}{2T} \sum_{j=1}^{T-1} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega_j) \delta(\omega_j)]^2 \right\} + o_p(1), \end{aligned}$$

where the equality follows from the law of large numbers. Combining this result with (B.17):

$$\begin{aligned} L_h(\phi_T) - L_f(\theta_0) &= \frac{1}{2T^{1/2}} \sum_{j=1}^T \text{tr} \left\{ f_{\theta_0}^{-1}(\omega_j) \delta(\omega_j) f_{\theta_0}^{-1}(\omega_j) (I(\omega_j) - f_{\theta_0}(\omega_j)) \right\} \quad (\text{B.19}) \\ &\quad - \frac{1}{8\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega + o_p(1). \end{aligned}$$

In the time domain:

$$\mathcal{L}_h(\phi_T) - \mathcal{L}_f(\theta_0) = -\frac{1}{2} \log \det(\Omega_f^{-1} \Omega_h) + \frac{1}{2} Y'(\Omega_f^{-1} - \Omega_h^{-1}) Y. \quad (\text{B.20})$$

Let

$$\Omega_\delta = T^{1/2}(\Omega_h - \Omega_f). \quad (\text{B.21})$$

The matrix $T^{-1/2} \Omega_f^{-1} \Omega_\delta$ satisfies the conditions in Lemma A1.1 in Dzhaparidze (1986). Therefore, the first term in (B.20) equals

$$-\frac{1}{2} \text{tr} \left\{ T^{-1/2} \Omega_f^{-1} \Omega_\delta - (2T)^{-1} [\Omega_f^{-1} \Omega_\delta]^2 \right\} + o(1).$$

By Lemma S.2 (i)-(ii) in the supplementary appendix, the leading term of the preceding display further equals

$$-\frac{1}{2T^{1/2}} \sum_{j=1}^T \text{tr} \left\{ f_{\theta_0}^{-1}(\omega_j) \delta(\omega_j) \right\} + \frac{1}{8\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega + o(1). \quad (\text{B.22})$$

The second term in (B.20) equals

$$\frac{1}{2T^{1/2}} Y'(\Omega_f^{-1} \Omega_\delta \Omega_h^{-1}) Y = \frac{1}{2T^{1/2}} Y'(\Omega_f^{-1} \Omega_\delta \Omega_f^{-1}) Y - \frac{1}{2T} Y'([\Omega_f^{-1} \Omega_\delta]^2 \Omega_h^{-1}) Y.$$

By Lemma S.2 (iii)-(v), the right hand side equals

$$\frac{1}{2T^{1/2}} \sum_{j=1}^T \text{tr} \left\{ f_{\theta_0}^{-1}(\omega_j) \delta(\omega_j) f_{\theta_0}^{-1}(\omega_j) I(\omega_j) \right\} - \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \right\} d\omega + o(1). \quad (\text{B.23})$$

Taking the sum of the leading terms in (B.22) and (B.23) leads to that of (B.19). Therefore, the time and frequency domain log likelihood ratios are asymptotically equivalent.

To simplify notation, let $A = (4\pi)^{-1} \int_{-\pi}^{\pi} \text{tr} \{ [f_{\theta_0}^{-1}(\omega) \delta(\omega)]^2 \} d\omega$. Under the null hypothesis of $f_{\theta_0}(\omega)$ being the true spectral density, the first term in (B.19) is asymptotically distributed as $N(0, A)$. Under the alternative hypothesis of $h_{\phi_T}(\omega)$ being the true spectral density, this term is asymptotically distributed as $N(A, A)$. Consequently, (B.19) is asymptotically distributed as $N(-\frac{1}{2}A, A)$ and $N(\frac{1}{2}A, A)$ under the two hypotheses, respectively. From these two distributions, it follows that the asymptotic power function is given by $\Pr(Z > z_{1-\alpha} - \sqrt{A})$.

Proof of Corollary 7. The proof is similar to that of Theorem 3. First, the result (B.4) continues to hold when the summations and integrations are taken over subsets of frequencies. Second, analogously to (B.5), we have

$$\frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} W(\omega_j) \text{vec} \left\{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \right\}^* \text{vec} \{ I(\omega_j) - f_{\theta_0}(\omega_j) \}. \quad (\text{B.24})$$

Relative to (B.5), here a complication arises because the Cesaro sum of the Fourier series of $W(\omega_j) \text{vec} \{ f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j) \}$ no longer converges uniformly over $[0, \pi]$ due to the discontinuities introduced by $W(\omega_j)$, i.e., the Gibbs phenomenon arises. However, their overall effect on (B.24) is asymptotically negligible following from the proofs in Qu (2014, Supplementary Material, the analysis of Display (B.4) in p.2). The rest of the proof is the same as the full spectrum case.

Proof of Corollary 8. We focus on the asymptotic distribution when μ_{θ_0} and $f_{\theta_0}(\omega)$ are the true mean and spectrum. We have

$$\begin{aligned} & \frac{1}{T} (\bar{L}_h(\phi_0) - \bar{L}_f(\theta_0)) \\ = & \frac{1}{T} (L_h(\phi_0) - L_f(\theta_0)) \\ & + \left(\frac{1}{4\pi} \right) (\mu_{\theta_0} - \nu_{\phi_0})' h_{\phi_0}^{-1}(0) (\mu_{\theta_0} - \nu_{\phi_0})' \\ & - \frac{1}{2T} \left\{ \log \det (h_{\phi_0}(0)) + \text{tr}(h_{\phi_0}^{-1}(0) I_{\phi_0}(0)) - \left(\frac{T}{2\pi} \right) (\mu_{\theta_0} - \nu_{\phi_0})' h_{\phi_0}^{-1}(0) (\mu_{\theta_0} - \nu_{\phi_0})' \right\} \\ & + \frac{1}{2T} \{ \log \det (f_{\theta_0}(0)) + \text{tr}(f_{\theta_0}^{-1}(0) I_{\theta_0}(0)) \}. \end{aligned}$$

The first term on the right hand side is already analyzed in Theorem 3. The second term gives rise to the additional term in $\bar{K}\bar{L}_{fh}(\theta_0, \phi_0)$ relative to $KL_{fh}(\theta_0, \phi_0)$. The third term, after multiplying by $T^{1/2}$, converges to a multivariate Normal distribution with mean zero and variance $(\mu_{\theta_0} - \nu_{\phi_0})' h_{\phi_0}^{-1}(0) (\mu_{\theta_0} - \nu_{\phi_0}) / (2\pi)$. The fourth term is $O_p(T^{-1})$. Finally, the first and the third terms are asymptotically independent because $E(\tilde{\varepsilon}_{ta} \tilde{\varepsilon}_{sb} \tilde{\varepsilon}_{uc}) = 0$.

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 model is estimated with the dataset from Smets and Wouters (2007) using the frequency domain quasi-Bayesian procedure based on the second order properties suggested in Qu and Tkachenko (2012). The two sample periods correspond to 1960:I-1979:II and 1982:IV-1997:IV. The prior distributions are taken from Table 1 in Lubik and Schorfheide (2004). In particular, the Inverse Gamma priors are $p(\sigma|v, s) \propto \sigma^{-v-1} e^{-vs^2/2\sigma^2}$, where $v = 4$ and s equals 0.25, 0.3, 0.6 and 0.2 respectively. When maximizing the log-posterior, β is reparameterized as $r^* = 400(1/\beta - 1)$ with r^* representing the annualized steady state real interest rate. The posterior distributions are based on 500000 draws using the Random Walk Metropolis algorithm with the scaled inverse negative Hessian at the posterior mode as the proposal covariance matrix. The first 100000 draws are discarded.

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 | 2.41 | 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 | 0.62 | 0.68 | 0.70 | 0.61 | 0.50 | 0.33 | 0.62 | 0.61 | 0.61 |
| ψ_2 | 0.23 | 0.27 | 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 | 0.61 | 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 | 0.58 | 0.57 | 0.56 | 0.58 | 0.59 | 0.60 | 0.59 | 0.62 | 0.61 |
| σ_z | 0.62 | 0.62 | 0.65 | 0.69 | 0.63 | 0.58 | 0.50 | 0.52 | 0.33 | 0.21 |
| $M_{r\epsilon}$ | 0.53 | 0.52 | 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 | 0.18 | 0.19 | 0.18 | 0.19 | 0.19 | 0.18 | 0.18 | 0.07 | 0.001 |

Note. KL denotes $KL_{ff}(\theta_0, \theta_c)$ with θ_0 corresponding to the benchmark specification. The 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 | 5.19E-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 | 0.0510 | 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 denotes $KL_{ff}(\theta_0, \theta_c)$ 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. Parameter values under alternative monetary policy rules, AS (2007) model

| | (a) Expected inflation rule | | (b) Output growth rule | |
|-------------------|-----------------------------|-------------|------------------------|-------------|
| | Indeterminacy | Determinacy | Indeterminacy | Determinacy |
| τ | 2.51 | 2.24 | 2.80 | 2.17 |
| β | 0.995 | 0.995 | 0.997 | 0.999 |
| κ | 0.49 | 0.84 | 0.48 | 0.82 |
| ψ_1 | 0.63 | 2.30 | 0.69 | 2.37 |
| ψ_2 | 0.54 | 2.25 | 0.01 | 0.01 |
| ρ_r | 0.87 | 0.65 | 0.86 | 0.64 |
| ρ_g | 0.66 | 0.93 | 0.66 | 0.93 |
| ρ_z | 0.60 | 0.88 | 0.59 | 0.88 |
| σ_r | 0.27 | 0.23 | 0.27 | 0.22 |
| σ_g | 0.58 | 0.77 | 0.57 | 0.77 |
| σ_z | 0.62 | 0.26 | 0.63 | 0.26 |
| $M_{r\epsilon}$ | 0.53 | – | 0.55 | – |
| $M_{g\epsilon}$ | -0.06 | – | -0.07 | – |
| $M_{z\epsilon}$ | 0.26 | – | 0.27 | – |
| σ_ϵ | 0.19 | – | 0.19 | – |

Note. The values are the minimizers of $KL_{fh}(\theta_0, \phi)$ under indeterminacy and $KL_{fh}(\theta_0^D, \phi^D)$ under determinacy, where h , ϕ and ϕ^D denote the spectral density and structural parameter vectors of the alternative models.

Table 5. 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 6. 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 | Investment 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 | Investment 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. |
| $\bar{\gamma}$ | Trend growth rate: real GDP, Infl., Wages |
| $100(\beta^{-1}-1)$ | Discount rate |

Table 7. Parameter values miminizing the KL criterion, SW(2007) model

| | θ_0^D | (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 |
| 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 | 0.85 | 0.85 |
| μ_p | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 | 0.68 |
| α | 0.19 | 0.19 | 0.19 | 0.19 | 0.19 | 0.19 | 0.19 |
| ψ | 0.54 | 0.54 | 0.54 | 0.53 | 0.54 | 0.53 | 0.52 |
| 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 | 1.35 | 1.33 |
| λ | 0.71 | 0.71 | 0.72 | 0.72 | 0.71 | 0.72 | 0.73 |
| ϕ_p | 1.60 | 1.60 | 1.61 | 1.61 | 1.60 | 1.59 | 1.59 |
| ι_w | 0.58 | 0.58 | 0.58 | 0.57 | 0.58 | 0.58 | 0.58 |
| ξ_w | 0.70 | 0.70 | 0.70 | 0.71 | 0.71 | 0.72 | 0.74 |
| ι_p | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 |
| 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 | 2.33 | 2.83 |
| r_π | 2.04 | 2.04 | 2.02 | 2.01 | 2.04 | 2.03 | 2.02 |
| $r_{\Delta y}$ | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.21 | 0.20 |
| 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 | 0.22 | 0.21 |
| 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 | 0.71 |
| ρ_r | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.14 | 0.14 |
| ρ_p | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 |
| 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 | 0.45 |
| σ_b | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 |
| 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 | 0.24 |
| 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 |
| $\bar{\gamma}$ | 0.43 | 0.43 | 0.43 | 0.42 | 0.45 | 0.50 | 0.56 |
| $100(\beta^{-1} - 1)$ | 0.16 | 0.17 | 0.19 | 0.22 | 0.15 | 0.11 | 0.07 |

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 8. 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 | 5.87E-04 | 1.88E-03 |
| T=80 | 0.0539 | 0.0706 | 0.0941 | 0.0574 | 0.0909 | 0.1384 |
| T=150 | 0.0553 | 0.0796 | 0.1159 | 0.0603 | 0.1108 | 0.1879 |
| T=200 | 0.0562 | 0.0852 | 0.1299 | 0.0621 | 0.1235 | 0.2206 |
| T=1000 | 0.0646 | 0.1505 | 0.3123 | 0.0800 | 0.2880 | 0.6169 |

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 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. 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 |

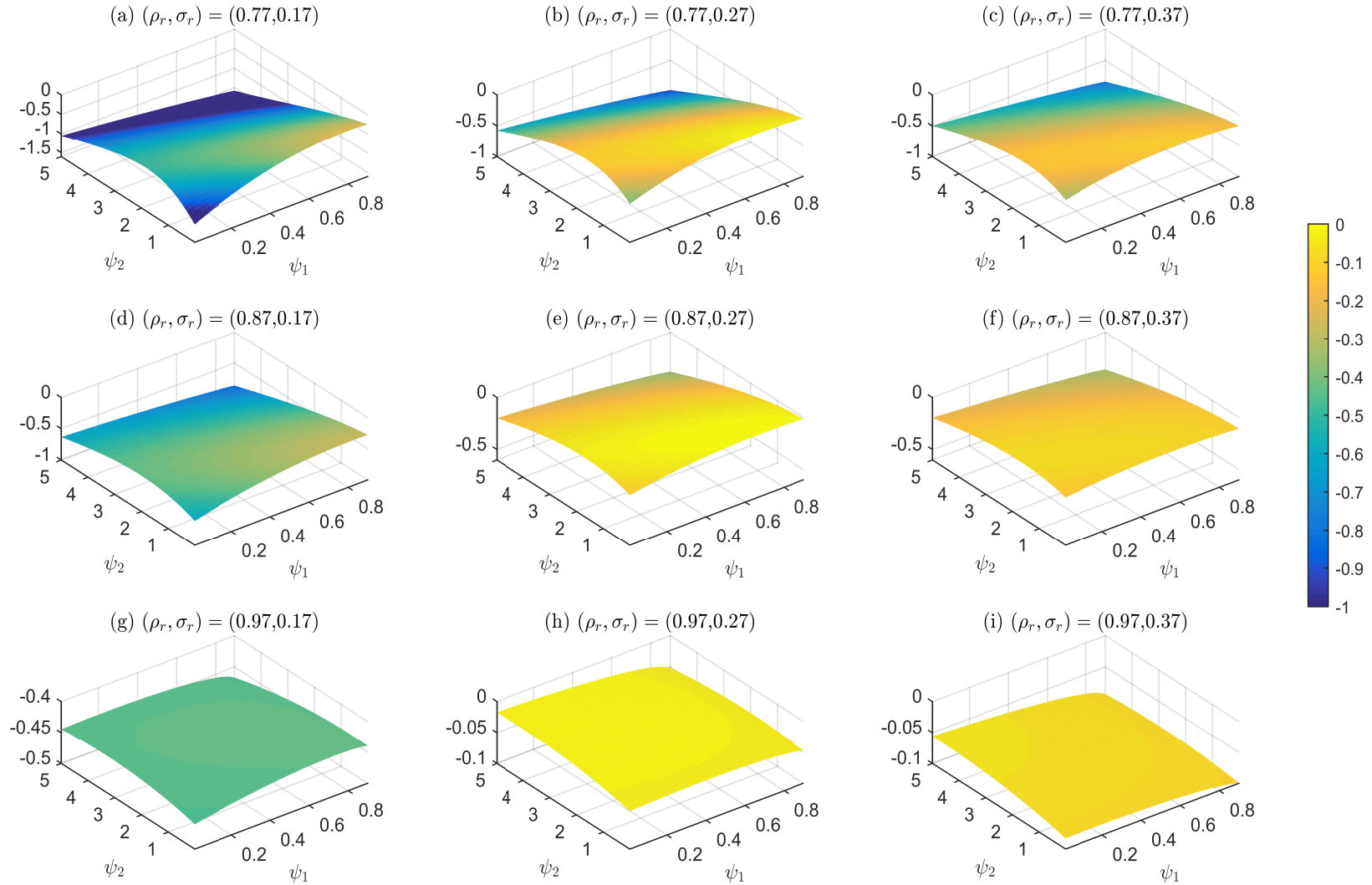
Note. KL and the empirical distance measure are defined as $KL_{fh}(\theta_0^D, \phi^D)$ and $p_{fh}(\theta_0^D, \phi^D, 0.05, T)$ with h and ϕ^D being the spectral density and structural parameter vector of the alternative model and T specified in the last four rows of the Table.

Table 10. The closest models with constrained real and nominal frictions, SW (2007) model

| | θ_0^D | $\xi_p=0.1$ | $\xi_w=0.1$ | $\iota_p=0.01$ | $\iota_w=0.01$ | $\varphi=1.0$ | $\lambda=0.1$ | $\psi=0.99$ | $\phi_p=1.1$ |
|-----------------------|--------------|-------------|-------------|----------------|----------------|---------------|---------------|-------------|--------------|
| 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 |
| $\bar{\gamma}$ | 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 |

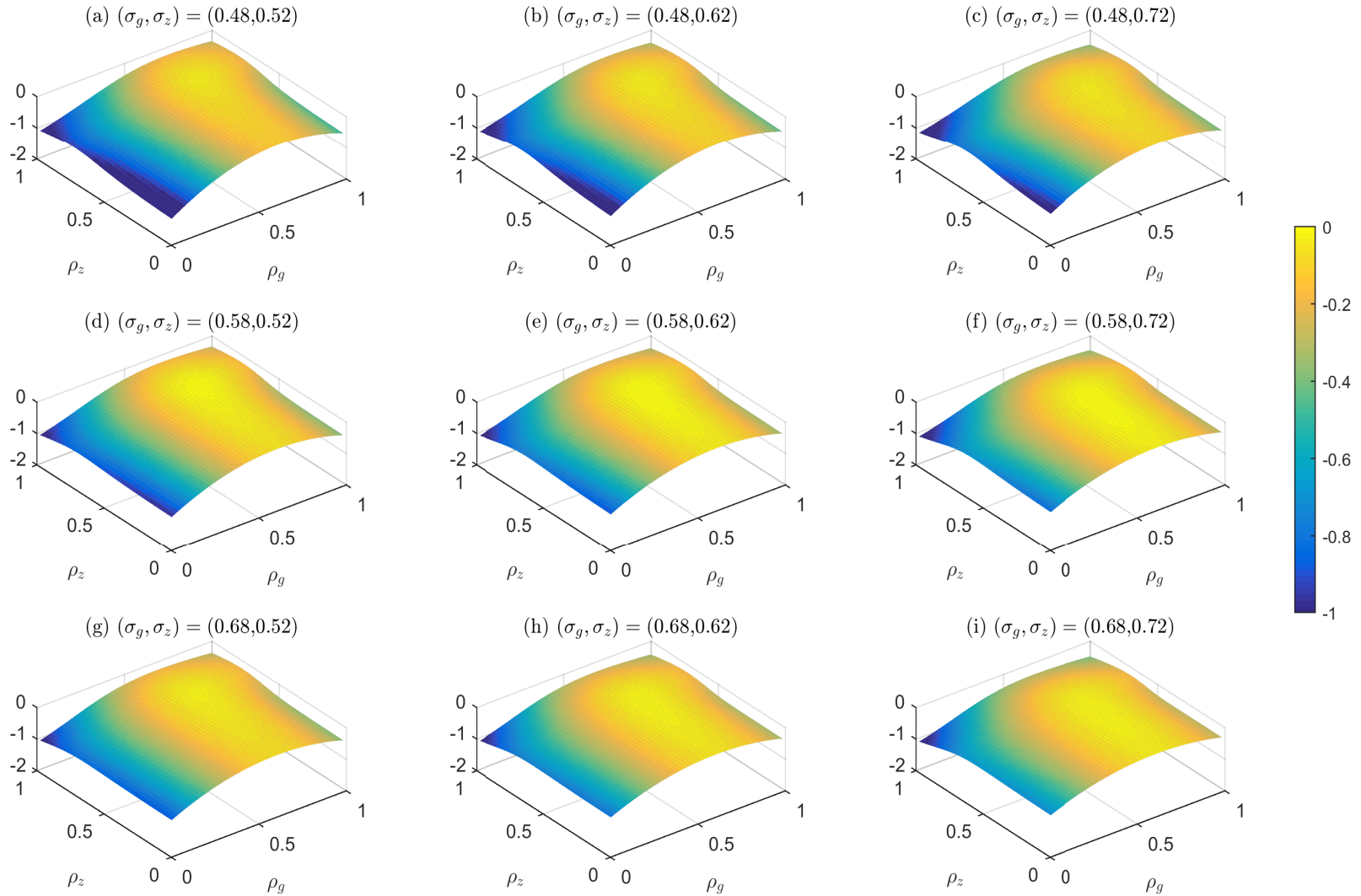
Note. KL (the second row) and the empirical distance measure (the third and fourth row) are defined as $KL_{ff}(\theta_0^D, \theta^D)$ and $p_{ff}(\theta_0^D, \theta^D, 0.05, T)$. Each column contains a parameter vector that minimizes the KL criterion under a particular constraint on the friction.

Figure 1. KL values when changing the monetary policy parameters



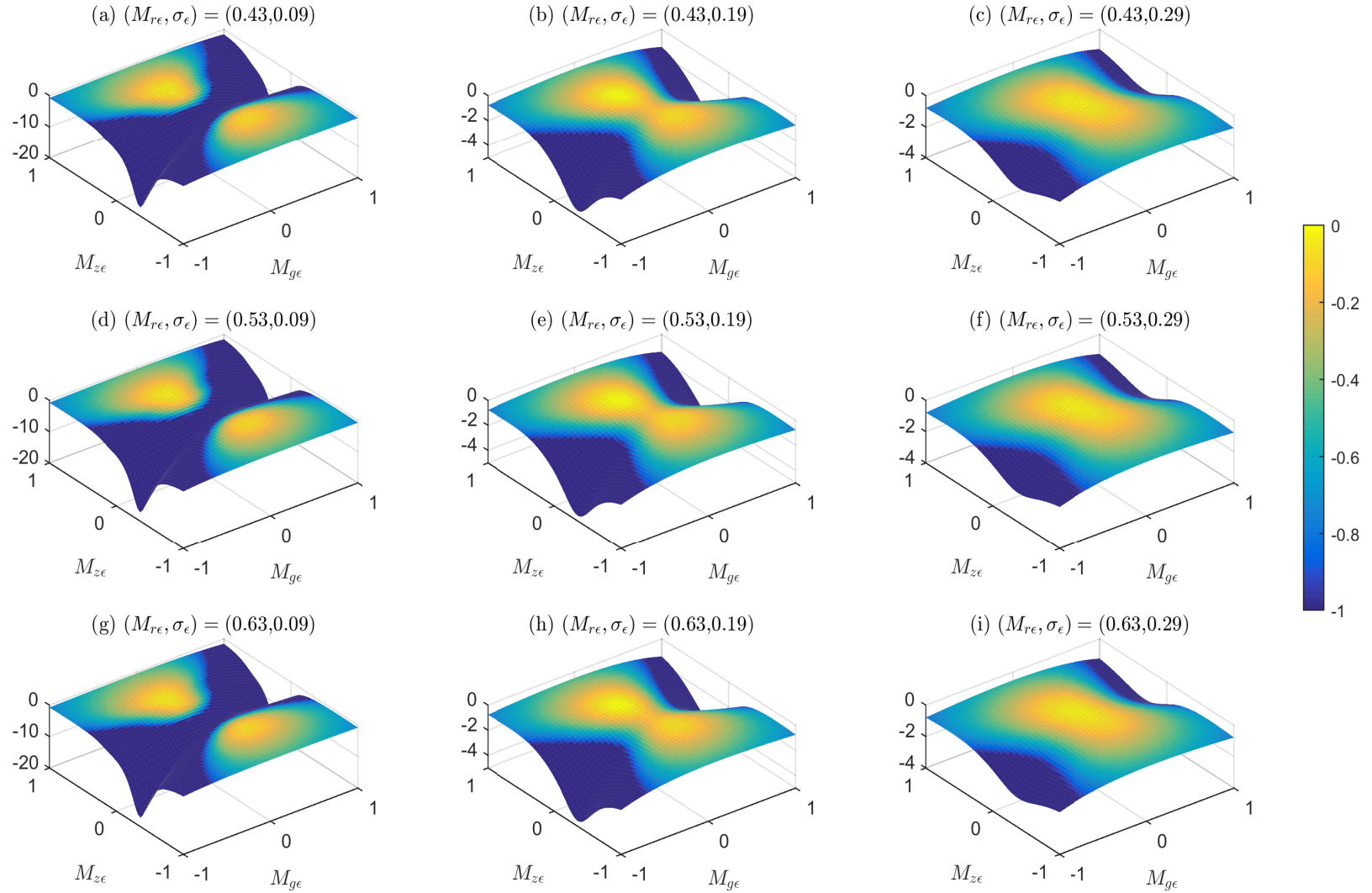
Note. Z-axis: the negative of KL; $\psi_1 \in [0.01, 0.9]$ and $\psi_2 \in [0.01, 5]$; ρ_r and ρ_g are equal to their values in θ_0 in Plot (e) and are increased or decreased by 0.1 in the remaining plots. The same color map is used throughout.

Figure 2. KL values when changing the structural shocks parameters



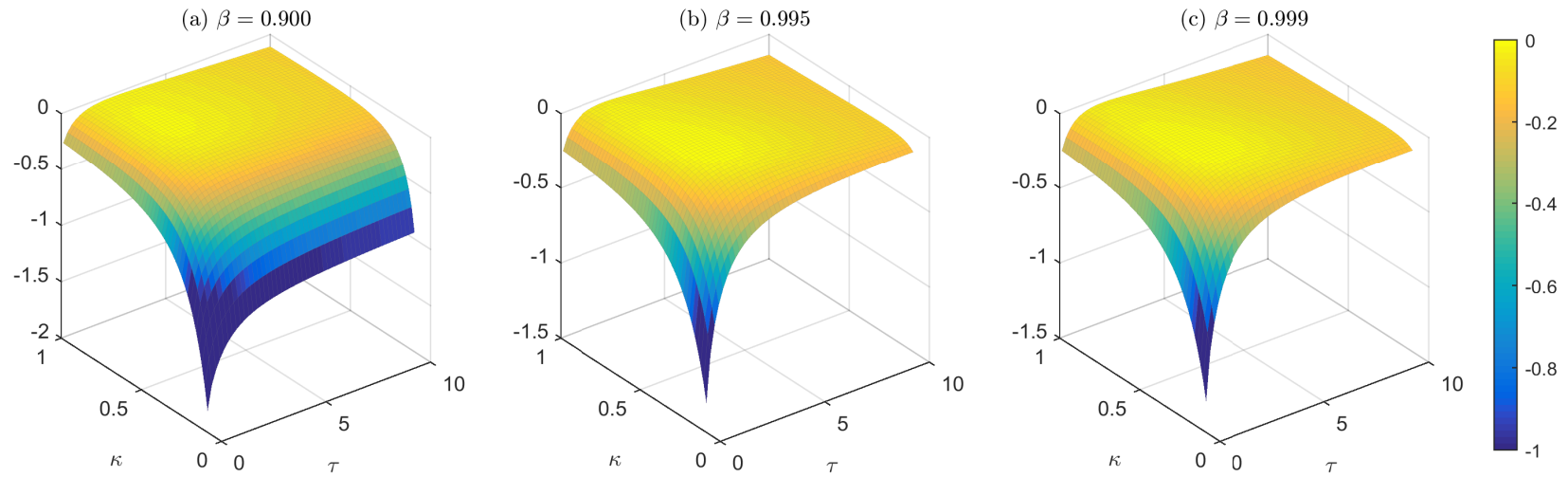
Note. Z-axis: the negative of KL; $\rho_z \in [0.01, 0.99]$ and $\rho_g \in [0.01, 0.99]$; σ_g and σ_z are equal to their values in θ_0 in Plot (e) and are increased or decreased by 0.1 in the remaining plots. The same color map is used throughout.

Figure 3. KL values when changing the sunspot parameters



Note. Z-axis: the negative of KL; $M_{g\epsilon} \in [-1, 1]$ and $M_{z\epsilon} \in [-1, 1]$; M_{re} and σ_{ϵ} are equal to their values in θ_0 in Plot (e) and are increased or decreased by 0.1 in the remaining plots. The same color map is used throughout.

Figure 4. KL values when changing the behavioral parameters



Note. Z-axis: the negative of KL; $\kappa \in [0.1, 1]$ and $\tau \in [0.1, 10]$; β is equal to its value in θ_0 in Plot (b). The same color map is used throughout.